351[illegible]89

AD A001766

1

Policy Capturing with Local Models:

The Application of the AID Technique in Modeling Judgment

L. L. Gooch, Major, USAF

Block #20

The past decade has witnessed increased research into the use of linear statistical models to describe subjective human judgment processes. Past studies have focused on, 1) the superior predictive capability of statistical models, vis-a-vis, man himself, 2) the detection and modeling of configural judgment process, and, 3) the use of judgment models as catalysts to conflict resolution. This research has generally been conducted in the context of clinical experiments and the results have not been extensively applied to practical decision situations. This lack of diffusion is primarily the result of failure of researchers to present their models in a format that the non-technical decision maker can understand and appreciate.

This dissertation helps to bridge the gap between the laboratory and operational environments by applying Policy Capturing to the decision process of an installment loan officer who approves or denies credit on the basis of a written application. The Automatic Interaction Detection (AID) algorithm was used to present the "structural image" of the policy in a format that is readily comprehensible by the decision maker. A unique conceptualization of configural decision processes is proposed and demonstrated. Experimental results given.

The dissertation is dedicated to

Penny,
Jennifer,
Winnie,
and Laura,

the four individuals who gave more to
this effort than I did.

Acknowledgments

John Donne's words, "No man is an island, entire of itself", were never more appropriate than when used to acknowledge the many people whose help and support made this research possible. Outstanding among these supporters are those who sponsored, supervised, and participated in this research effort.

I would like to first express my appreciation to the Air Force Institute of Technology for having made my entire graduate education possible. Next, I would like to recognize the considerable contribution of the members of the Operations Research and Applied Statistics Group at The University of Texas at Austin. It was their guidance, instruction, and motivation that allowed me to exploit the opportunity, offered by AFIT, to undertake this effort.

Throughout the research itself, the unwavering support and guidance of my supervising professor, Dr. G. R. Wagner, has been a continuing source of inspiration and will be forever appreciated. The additional encouragement and timely advice offered by the other members of my committee, Dr. F. Brons, Dr. E. Jennings, Dr. F. May, and Dr. J. H. Ward, Jr. undoubtedly made this research a more meaningful experience. It is indeed an honor to have been supervised by such an august group of educators.

My special appreciation must be expressed to Mr. Dave Statham, and his staff at Lamar Savings and Loan Association for their enthusiastic participation in this project. Their patient response to my continual requests for "more data" will hopefully reap tangible dividends in the near future.

The considerable assistance of the Personnel Research Division, Air Force Human Resources Laboratory, Lackland, AFB, Texas must be acknowledged. Their own pioneering efforts in the field

of Policy Capturing and their generous support of my efforts to advance the frontier of knowledge in the field were of inestimable value in the completion of this research.

Finally, the individual efforts of the many other people who contributed both directly and indirectly will long be remembered as I reflect upon this experience. To them, I simply say, THANKS!

L. L. G.

The University of Texas at Austin
December, 1972

ABSTRACT

The use of linear statistical models to describe subjective judgment processes has received considerable attention in the fields of clinical psychology and psychometrics over the last decade. The majority of this research has been conducted within the basic conceptual framework of the Brunswikian Lens. The studies have focused on three areas. They are: 1) the superior predictive capability of the statistical models, vis-a-vis, man himself, 2) the detection and modeling of configural judgment processes, and, 3) the use of judgment models as catalysts to conflict resolution. This research has generally been conducted in the context of clinical experiments with only limited implementation of the results in the business, academic, and administrative environments. The lack of diffusion of this innovative concept is primarily attributable to the failure of the researchers to present their models in a format that the non-technical decision maker can understand and appreciate.

This dissertation attempts to help bridge the gap between the laboratory and operational environments by applying the techniques of Policy Capturing to a practical decision process. The judgment process used is that of the installment loan officer who approves or denies credit solely on the basis of information contained on a written credit application and the past credit rating of the applicant. The Automatic Interaction Detection (AID) algorithm is used as the primary means of describing the loan officers' policies. It presents the "structural image" of the policy in a graphical format that is readily comprehensible by the decision maker. The analytical data provided by the technique is used to hypothesize the appropriate mathematical model.

A unique conceptualization of configural decision processes as a series of local models over subspaces of the predictor set is proposed and demonstrated. A heuristic procedure for defining local models is presented and used to develop policy models for the loan officers. These models achieve cross-validation hit rates in excess of 91 percent in predicting actual decisions on new loan applications. Models for identifying the top 30 percent of the applications, in the absence of credit rating data, are developed and discussed as a means of improving the level of service that the lending institution can provide during non-banking hours.

TABLE OF CONTENTS

CHAPTER		Page
VI.	ANALYSIS, RESULTS, AND IMPLICATIONS OF LOAN OFFICER POLICY MODELS	

LIST OF TABLES

Table		Page

LIST OF FIGURES

Figure		Page

Figure		Page

INTRODUCTION

The many definitions of Operations Research all involve a common conceptual thread; that is, "the application of the scientific approach and mathematical techniques to describe, control, and improve operations of organizational systems". This dissertation represents an effort to adapt, expand, and introduce yet another technique into the inventory of tools available to the practitioner of Operations Research.

The voluminous literature in the field of Operations Research is replete with decision models and solution algorithms for problems in which both the objective functions and the relevant parameters are thought to be well known. The general statement of such problems is:

OPTIMIZE:	$F(\tilde{x})$	$\tilde{x} \in V^n$
Subject to:		
equality constraints:	$h_i(\tilde{x}) = 0$	$i = 1 \ldots m$
inequality constraints:	$g_i(\tilde{x}) \leq 0$	$i = m + 1 \ldots p$

In all cases the coefficients of the objective function and constraints are either deterministic or, at least, estimable by some probability distribution. In those cases where the objective function, $F(\tilde{x})$, is not known, it is either assumed, or the problem is left unsolved.

The lack of objective functions has spurred research in such fields as "Utility Theory" (Fishburn, 1968) and "Measurement Theory" (Roberts, 1970), and has brought about such techniques as Delphi (Helmer, 1967) (Dalkey, 1971) and the Echo Method (Milburn, 1968). However, these techniques have not really been refined to the stage

that they represent readily usable tools for the practitioner of Operations Research.

The technique proposed here represents a more direct attack on solving the problem of determining what is important to the decision maker and provides a readily-usable methodology for analyzing, describing, simulating, and possibly improving the decision-making function. This approach involves the inference of the values of the decision-maker through the analysis of his decisions over a sufficiently large number of sample decision situations. Development of this type of procedure was advocated by Churchman (1961) in his philosophical discussion of optimal decision processes.

The research chronicled by this dissertation includes the exploration and adaptation of a concept known as "Policy Capturing" for use in ecologically valid environments. The basic concept of "Policy Capturing" has been an area of continued research in the field of clinical psychology during the past decade. This research has been pursued and results reported under such titles as "Judgment Modeling", "Decision Modeling", and "Analysis of Cognitive Processes". Although there has been much work done and many articles written in the various journals of clinical psychology, the transition from a clinical experiment to an applications methodology has only been attempted in a limited number of instances. The lack of diffusion of this innovative concept is mainly due to the approach taken by many researchers of working with contrived and synthetic situations, a practice which has not enhanced the adaptability or implementability of the scientific findings of the psychologist to the extra-laboratory or practical decision situation.

It is the purpose of this research to broaden the applicability of the Policy Capturing concept by bridging the gap between the

findings of the clinical psychologists and the realities of typical operational decision situations facing the Operations Research practitioner. This bridge is afforded by the adaptation of recently developed satistical techniques for the purpose of defining policy models that reflect the underlying judgment process of the decision-maker and retain a high degree of predictive stability.

This dissertation is divided into seven chapters and three appendices. Chapter 1 presents the basic framework and the experimental paradigm within which the Policy Capturing concept is formulated and the current research was performed. Chapter 2 provides a review and analysis of the past research that is relevant to the current efforts and Chapter 3 provides the focus and motivations for the current research. Chapter 4 describes the development, modification, and interpretation of the statistical techniques used in this research. This chapter also presents a methodology for developing "structural image" models that provide the primary bridge between the concept of Policy Capturing and the practical methodology for its application. Chapter 5 presents the background and case history of the practical environment within which this research was conducted and the viability of the methodology is demonstrated. Chapter 6 presents results, conclusions and implications generated by this research for the particular practical environment in which it was used. Chapter 7 relates the current research to the findings and hypotheses generated by various previous researchers and discusses the implications of applying the basic methodology to other environments.

Finally, three appendices are provided. Appendix A reflects a comparison of the various statistical tools available and their relationship with the Policy Capturing methodology. Appendix B

provides the forms, data, and detailed structural-image models for the case study. Appendix C provides a self-contained description for using the computer programs and the analysis scheme, developed in this research, in future Policy Capturing efforts.

Guide to the Reader:

Whereas this dissertation attempts to straddle the gap between the technical intricacies of statistical modeling and the operational concerns of manager, it necessarily addresses itself to two audiences. A guide to those sections of interest to each seems in order. The reader who is interested in the details of HOW the analysis was performed and the models were built should concern himself with Chapters 1-4, 7, and Appendices A and C. The reader who is interested mainly in the results of this application of Policy Capturing to modeling the policies of loan officers will be most interested in Chapters 1, 5-7, and Appendix B.

CHAPTER I: CONCEPT OVERVIEW

The Judgment Modeling Concept:

The judgment process is defined as "the process of forming an opinion or evaluation by discerning and comparing". Implicit in this definition is the existence of one or more attributes or characteristics of the situation that need to be compared or evaluated, and the identification of these attributes. If one makes many judgments of the same nature, the logical premise of consistency would dictate that the same set of evaluations and comparisons should be carried out in each decision situation. Such a set of comparisons could be classified as a model or policy for making all judgments of a particular nature. If the attributes can be quantified in some mathematical form, the comparisons and evaluations can be carried out mathematically with the policy being embodied in the various mathematical operations to be performed.

The fundamental premise of the Judgment Modeling Concept is that it is possible to represent subjective human judgment with objective mathematical models. This premise, in itself, has been a source of considerable debate relative to the superiority of clinical vs. actuarial decisions. This debate was initiated when Meehl (1954) proposed that psychologists could make better diagnoses, on the average, if they were to actuarily combine the same data that they normally used in their clinical diagnosis. The ensuing research has resulted in over 600 studies in the rather narrow field of human information processing in the last decade. (Slovic and Lictenstein, 1970)

The Judgment Modeling Process:

The definition of judgment modeling itself implies a rather straight-forward, logical sequence of events that has been the basis for most of thr research in the field. The functional elements of this process are shown in Figure 1-1. As might be expected, each element has been the focai point for much detailed research, with particular attention being paid to the formulation of the mathematical models in step number three and to the comparison of judgment models in step number four. Very little research has been accomplished on the process as a whole and virtually none has been attempted on the process in practical, extra-laboratory environments.

In an excellent review of past research, Slovic and Lictenstein (1970), have reviewed and compared the Regression and the Bayesian Approaches of judgment modeling. The basic differences between these two approaches being that in the former, the judgment process is formulated in the context of the general linear hypothesis while in the latter, the judgment process is formulated in the context of conditional probabilities and Bayes' Theorem.

The work pursued in the research discussed herein has all been within the Regression Approach and, therefore, the Bayesian Approach will not be addressed. Further, within the Regression Approach, the main thrust of this research has been within the "correlational paradigm" with some efforts encroaching on the area traditionally known as the "functional measurement paradigm". The difference between these paradigms rests essentially in the elements of the judgment process considered, the statistical techniques used, and the goals pursued within a particular research effort.

FUNCTIONAL ELEMENTS OF THE JUDGMENT MODELING PROCESS

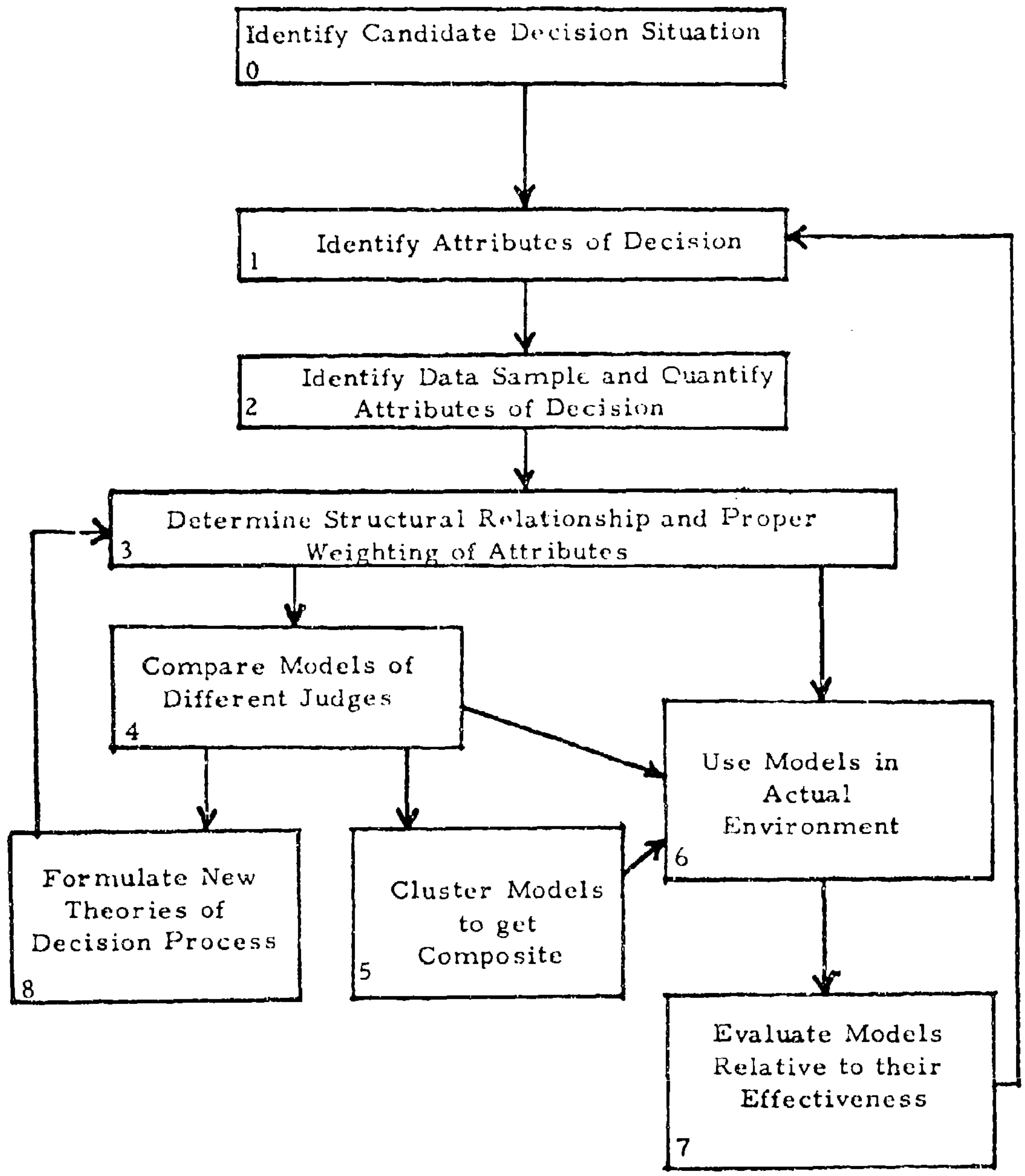


FIGURE 1-1

The Brunswikian Lens:

The conceptual model for the Regression Approach of judgment modeling was first proposed by Egon Brunswik in 1952. The model as shown in Figure 1-2, or minor modifications thereof, has served as the conceptual framework for both the correlational and functional measurement paradigms.

The Lens model graphically depicts the underlying hypothesis that each attribute (cue) of the decision situation is probabilistically related to an ecological criterion of a "true state-of-nature". This relationship can be characterized by a general linear model of the form:

$$Y_e = \hat{Y}_e + \epsilon_e = \sum_i B_{i,e} X_i + \epsilon_e$$

where:

Y_e = the actual criterion value

$\hat{Y}_e$ = the predicted criterion value

ϵ_e = the error in the prediction

$B_{i,e}$ = the coefficient of the i^{th} variable

X_i = the value of the i^{th} variable

e = the subscript representing the ecological side of the Lens

On the judgmental side of the Lens, a similar model is hypothesized which relates the judged value of the criterion to the

BRUNSWIKIAN LENS MODEL

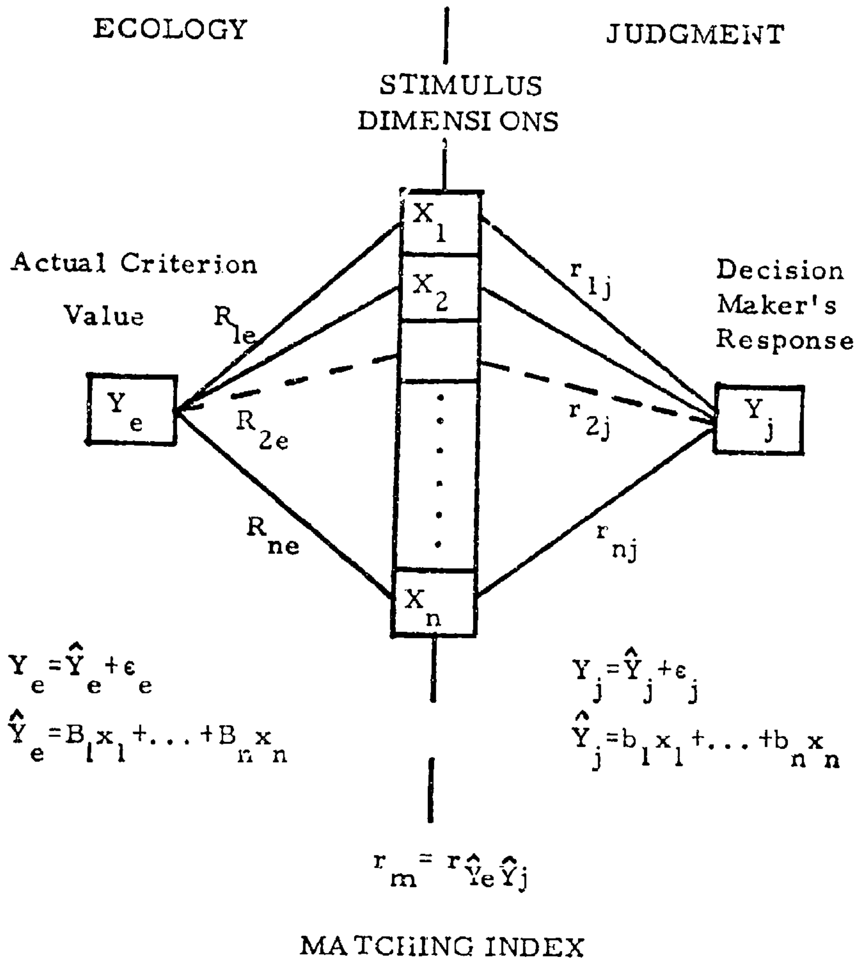


FIGURE 1-2

stimulus cue values. This equation is:

$$Y_j \quad = \quad Y_j + \epsilon_j \quad = \quad \sum_i b_{i,j} X_i + \epsilon_j$$

where:

Y_j	=	the judged value of the criterion
Y_j	=	the predicted value of the criterion per the judgment model
ϵ_j	=	the error in the prediction
b_{ij}	=	the coefficient of the i^{th} variable
X_i	=	the value of the i^{th} variable
j	=	the subscript representing the judgmental side of the Lens

Further statistics that are relevant to using and discussing the model are:

$R_{i,e}$	=	ecological validity: correlation of the i^{th} cue to the true criterion value (Y_e)
$r_{i,j}$	=	utilization coefficient: correlation of the i^{th} cue to the judged criterion value (Y_j)
r_e	=	$r_{Y_e Y_e}$: multiple correlation between the actual and predicted criterion value in nature
r_j	=	$r_{Y_j Y_j}$: response consistency: multiple correlation between decision makers judgments and the value of those judgments predicted by the model

$$r_a = r_{Y_e Y_j} : \text{ achievement index}$$

$$r_m = r_{\hat{Y}_e \hat{Y}_j} : \text{ matching index}$$

In essence, the left-hand side of the lens represents a modeling of reality and the right-hand side of the lens represents a modeling of the judge's perception and interpretation of reality. On the ecological side of the lens, the error, ϵ_e, is made up of random error and possibly any lack-of-fit of the hypothesized model to reality, whereas, on the judgmental side, ϵ_j, is composed of lack-of-fit of judge's model and his own inconsistency in following his model. It has been pointed out by Ward (1970) that the lack-of-fit portion of the error in the judgmental equation can be attributed to two main causes, 1) missing stimulus variables in the hypothesized relationship, and 2) incorrect structural form of the hypothesized variables in the model.

Correlational versus Functional Measurement Paradigm:

The distinction between the correlational and the functional measurement paradigms can be clarified in terms of the lens model. In the correlational paradigm, the emphasis is on the judgmental side of the lens with the structure of the judgmental model, the cue coefficients ($b_{i,j}$), and the response consistency (r_j) being of most interest. Elements 3, 4, 5, and 8 of the judgment modeling process have been the main areas of concern for researchers within this paradigm. In the functional measurement paradigm, interest is still evidenced in the judgmental side of the lens, but particular emphasis is put on the modification of the judgmental equation as a result of feedback from

success and failures of applying the model. Statistics of particular interest to workers within this paradigm are the achievement index (r_a) and the matching index (r_m), with most of the experimental work occurring within elements 1, 2, 3, 6, and 7 of the judgment modeling process.

Definition of Policy Capturing:

It should be noted here that in past research, the term "Policy Capturing" has been generally used to describe the analysis of the judgmental side of the Brunswikian Lens. However, for want of more descriptive terminology, the connotation of "Policy Capturing" has been expanded within this research to be as follows:

IDENTIFICATION AND QUANTIFICATION OF THE ATTRIBUTES THAT ARE PERTINENT TO A DECISION AND THE SUBSEQUENT MATHEMATICAL DESCRIPTION OF THE DECISION POLICY FOR THE EVALUATION OF THESE ATTRIBUTES is defined as POLICY CAPTURING.

Definitions:

At this point, it is beneficial to explicitly define various terminology that will be used throughout the description of past and current research in order to avoid any semantic difficulties that might arise relative to various types of mathematical models.

Linear Models: A linear model is of the form

$$Y = \hat{Y} + E = \sum_i b_i X_i + E \qquad i = 1 \ldots . n$$

where: $\hat{Y}$: $\sum_i b_i X_i$ = predicted value of the criterion

b_i: are parameters to be estimated

X_i: are variables that are known for any given data entity or that can be compounded from other basic variables such as $X_i = Z_i * W_i$ or $X_i - Z_i^3$

E: the error

The definition "linear" refers to the fact that the parameters "b_i" are constant over the entire predictor space (V^n) and does not relate to the order or power of the variables "X_i". All models considered as being linear models.

Main Effects Models: A linear model in which the highest power of any variable is unity.

Curvilinear Models: A linear model in which the power of some variable is greater than unity, i.e., X_i^3.

Configural Effects Models: A linear model in which cross-product terms of the form $b_{ij} X_i X_j$ where X_i and X_j are different variables, are included. This is also referred to as an "interaction" model.

Local Models: This is a linear model that applies over some subspace of the predictors, $S^m \epsilon V^n$, and can be considered as a special case of the configural model. Whereas in configural models, the terms $b_{ij} X_i X_j$ are generally considered to apply throughout the entire predictor space, V^n, in local models, such terms are applied only over the local subspace S^m.

Judges: This is the common term for the individual rater or decision-maker who is making the judgment. Many of the past works use the terminology "subjects".

<u>Cues</u>: Variables, attributes, predictors, or other characteristics of the decision situation which are perceived by the judge and are hypothesized as the stimuli for his decisions.

<u>Historical Data</u>: Cue and decision data taken from recorded files, that is, the cues were considered and the decisions made prior to the introduction of the concept of Policy Capturing to the subjects.

CHAPTER II: REVIEW AND ANALYSIS OF PAST RESEARCH

Overview:

The bulk of past research in the area of judgment modeling has been accomplished at three institutions, each with a somewhat separate focal point. At the Oregon Research Institute (ORI) the interest of such researchers as Hoffman, Slovic, Goldberg, and Dawes has been on elements 3, 4, and 5 of the judgment modeling process as shown in Figure 1-1. Their main thrust has been in evaluating how well a model-of-man can do versus man himself and in investigating the best way to formulate the judgment models.

At Colorado's Institute for Behavioral Science, the focus of researchers Hammond, Hursch, Todd, Earle, and Brehmer has been on elements 4 and 7 of the process. Their main efforts have been in investigation of policy conflicts, interpersonal learning, and the training of subjects to use various types of cues.

The only real experience in using judgment models in a practical environment to date has been at the Air Force Human Resources Laboratory (AFHRL) where researchers Ward, Christal, Koplyay, Madden, and Mead have focused mainly on elements 4, 5, 6 and 7 of the process. They have specialized in capturing and clustering the policies that individual board members use in making USAF personnel decisions.

Although these institutions have taken a major lead in the past research in Policy Capturing, independent researchers including Einhorn (elements 3 and 8), Green (element 3), and Kleinmuntz (element 1) have made substantial contributions to the current body of knowledge in the field.

The Early Regressionists:

In 1960, Hoffman reported that one could "paramorphically" represent the policy of a decision maker very well with first-order linear regression models. He applied the term "paramorphic" to denote the fact that, even though the individual terms in the policy model did not necessarily represent meaningful physical constructs relative to the judge's logical processes, these terms could be combined so as to predict a set of judgments (usually rankings) that had a high correlation with the actual judgments. Hoffman contended that even though these policy models were paramorphic and might not portray the decision process accurately from a structural standpoint, they could account for a significant amount of the variability in individual judgments. He felt that useful information could be gained about the judge's dependence on each cue by examination of the Relative Weight for each cue. He defined the Relative Weight for cue "i" in the policy of judge "j" as:

$$RW_{i,j} = \frac{\beta_{i,j} r_{i,j}}{r_j^2}$$

where:

$\beta_{i,j}$ = standardized regression weight of the i^{th} cue for the j^{th} judge's policy.

$r_{i,j}$ = correlation coefficient of the i^{th} cue with the criterion for the j^{th} judge. (Hoffman calls this the utilization coefficient.)

r_j = response consistency of the j^{th} judge.

This contention was quickly refuted by Ward (1962) in a demonstration that the Relative Weight was meaningless except in the very special case of orthogonality among the predictors.

Consequently, many researchers resorted to near total reliance on factorially designed experiments in order to have meaningful measures of comparison among predictor variables. To facilitate the comparison of policies on an inter-judge basis, a trend toward rather strict control of the data sets used in the analysis also developed. Hoffman (1960) justified such control by stating:

> ...restricting the situation (by controlling the stimuli) assures that each person is evaluated with respect to the same information. Ambiguous and equivocal cues are removed, and all judges are thereby certain to have at their disposal the same information and no more. The inferences made beyond this point are certain to have their origins in the data provided, (p. 118).

With his original work, Hoffman sparked two areas of controversy that are germain to the problem of applying Policy Capturing to an actual decision making environment. The first area is the use of first-order linear models as an adequate representation of the judgment process regardless of the underlying logical procedure. This approach can present a nearly insurmountable obstacle to practical implementation in convincing the decision maker, who may be unsophisticated mathematically, that an unexplainable equation can do as well as he can with his explainable judgment procedure. The second area is the use of highly structured and controlled data sets for capturing the policy of the decision maker. This technique often leads to a lack of credibility since the data used to capture the policy may not conform to the "real world" data that the decision maker must assess in his normal judgment process.

The adequacy of a paramorphic first-order model must really be evaluated relative to its intended use. If prediction of the final outcome is in the major purpose and there is no "selling to the management" problem involved, then a highly efficient first-order paramorphic is probably better than a more complex configural model.

This is exactly the situation confronting the research personnel at the USAF Human Resources Laboratory, Lackland AFB, Texas. These researchers were particularly concerned with providing decision aids to the numerous personnel boards which make such decisions as "Who should be promoted?" or "In which occupational specialty should a recruit with a given set of background qualifications be trained?" The goal in these instances is adequate prediction of the composite decisions of many judges for many cases and not the detailed analysis of how the individual judges make up their minds. The "selling job" is also minimized in that they are operating in an environment where they do not have to convince a given judge that the policy model is congruent with "the way he thinks he makes decisions", rather, the paramorphic policy model sells itself if it is predictively efficient in matching the composite decisions of the board members and can result in increased operational efficiency.

The AFHRL has published numerous studies in the development of policy models for specific areas. A few of these areas include assignment of personnel to jobs (Ward, 1962), merited pay and merited grade for jobs (Madden, 1963), promotion of enlisted men (Koplyay, 1969), and evaluation of job difficulty (Mead, 1970). In most of these studies, the key points of interest has been the comparison of the individual policies of the board members and the attainment of a composite policy model. Thus, they have emphasized the clustering and comparison of the policy equations and have developed and applied the JAN

Technique. With JAN, the captured policies of the individual judges are hierarchically clustered in a manner that minimizes the loss in predictive efficiency of the composite policies at each step. (Ward and Hook, 1961)(Bottenberg and Christal, 1961)(Christal, 1963)

An example of the information to be gained about the consistency of a given rater in applying "his policy" and about the homogeneity of "policies" within a group of raters can be illustrated by considering the results published by Madden. (1963) The data on Table shows the distribution of "rater efficiencies" (the multiple r_j^2 for the rater's paramorphic policy model) for each of 38 raters performing the ratings on each of 50 separate job-descriptor profiles. The "rater efficiency" reflects the ratio of the variation in ratings explained by the raters' policy model to the total variation in the ratings. It has a theoretical upper limit of unity. This ratio is an indication of the quality of the "fit" of the policy model to the policy used by the rater in making the ratings. Note that an efficiency of .99 was attained in modeling the policy of one individual relative to merited pay and this efficiency was attained for five individuals relative to merited grade.

Table 1. shows the results of hierarchial clustering of the policy models of the raters. At each stage of the hierarchial clustering procedure, the number of groups or clusters is reduced by one by combining that set of policy models that reduces the composite r_j^2 the least amount. A sudden drop in r_j^2 at a given stage would indicate that two non-homogeneous clusters had been grouped at that stage. Note that in the last step of the procedure for the merited pay policies, a group of 37 raters was combined with a lone rater and the resultant r_j^2 dropped from .87 to .78. By using clustering techniques, one can determine if significant differences in policy exist among individuals or groups of individuals on a board. Then action can be taken to eliminate

Table 1-1

RATER CONSISTENCY

Distribution of r_j^2's Achieved in Predicting Rank Order of Merited Pay and Merited Grade from Factor (predictor) Values

r_j^2	Merited Pay	Merited Grade
.99	1	5
.98	2	7
.97	6	2
.96	6	2
.95	6	2
.94	3	2
.93	1	0
.92	4	3
.91	1	0
.90	3	2
.89	1	5
.88	1	1
.87	1	2
.86	1	1
.85		1
.74	1	
.72		1
.31		1
.15		1
N=	38	

From: (Madden, 1963)

Table 2-2

POLICY COMPARISON

Effects on r_j^2 of Number of Groups

Basis of Ranking	Number of Groups	Number of Judges Each Group	Comp r^2
Pay	38	1	.93
	3	33, 4, 1	.88
	2	37, 1	.87
	1	38	.78
Grade	34	1	.92
	3	16, 16, 2	.88
	2	32, 2	.86
	1	34	.84
Pay & Grade	72	1	.93
	3	64, 5, 3	.86
	2	67, 5	.83
	1	72	.81

50 Subjects: 10 Predictors: No Interactions: 10 Levels per predictor

*Note that only 34 raters were grouped on the Merited Grade equation due to obvious inconsistency in 4 of the raters.

From: (Madden, 1963)

members with deviate policies or to resolve conflicts and differences of opinion among the raters.

In a AFHRL sponsored study of the effectiveness of identifying and resolving policy differences by first capturing the policies and then using the JAN Technique, Mullins and Usdin (1970) compared this procedure with one in which each judge estimated the relative importance of each variable and then the judges mediated their weights for each of the variables and came up with a composite equation. They found that by using the JAN Technique, the quality of the group judgment always improved but that by using the latter technique, the mediated weight for a variable was often only a reflection of the opinion of the most powerful, or the most vocal, judge. Further, they found the quality of the group estimate was worse as often as it was better.

The use of policy aggregation as a catalyst to conflict resolution was further pursued and reported on by Stephenson and Ward (1970) in an experiment wherein they captured the policy model of each of several members of a promotion board and clustered their individual policy models. This resulted in a focusing of attention on the previously non-isolateable reasons for differences between board members in the ratings of potential promotees. With areas of difference brought into focus by means of different weights on the relevant variables, some limited progress was reported in getting the board members to come to a common agreement on the relative importance of the various factors. They reported that even though compromise on a policy did not result to the desired degree, the judges in the study reported that the technique had served a very valuable purpose in clarifying the positions and policies of each of the board members.

Both of these studies would suggest that policy capturing could provide a solution to the problem in face-to-face communications described by Robinson (1971) as "noise"; that is, discussion of various

interests and not problem solving.

In most of the work carried on or sponsored by the AFHRL, a difference in policy between several judges, or groups of judges, has not been decomposed further than necessary to effect a resolution to a common policy or to identify and eliminate judges with greatly deviant policies. More recently, Brehmer (1970) has pointed out that apparent policy disagreements should really be decomposed into spearate components, one part due to individual policy inconsistency, and the other part due to inter-judge policy conflicts. His experimental results indicate that as judges with policy disagreements work to reduce the conflict that exists on an inter-judge basis, there is a counter-balancing effect with reduction in policy conflicts being accompanied by increases in individual policy inconsistency, resulting in very little net change in apparent differences. This would indicate that once people recognize the innate causes of policy conflicts, they may be willing to change their weightings of the cues, but they have a difficult time effecting the change because they cannot apply the new or "compromise" policy consistently. Brehmer also found that policy differences were reduced faster by free communication and free communication plus objective cue weights than they were by just communication of differences in judgments, a result that would support the utility of Policy Capturing as a catalyst for reducing policy differences.

The Configurality Debate:

A somewhat different set of interests has motivated the researchers for whom the first-order paramorphic model was untenable. They have persisted in what Goldberg (1968) has referred to as the "search for the configural judge". This continued effort to model the

policy of the judge with terms that are more reflective of the actual logic process of the judge is stimulated by: 1) the judges' introspective beliefs that they process cues in complex and configural ways, 2) the possibility that truly configural judgment tasks have not yet been studied, and, 3) the possibility that the experimental designs and statistical techniques available to the researchers were best suited for discovery of first-order effects and not particularly suited for unmasking configural relationships. (Slovic and Lictenstein, 1970) Many publications have appeared on both sides of the "first-order" versus "higher-order and configural" issue; the principle participants being Wiggins and Hoffman (1968), Goldberg (1968), Hoffman, Slovic and Rorer (1968), Einhorn (1970, 1971) and Valenzi (1970).

Evidence for the tremendous power of the first-order model was presented by Yntema and Torgerson (1961) when they demonstrated that 94% of the variance within a completely configural design could be accounted for by a first-order model.

Their analysis was based on synthetically generated data from a purely configural 3^7 design of the form:

$$W = aXZ + bXY + cYZ$$

with their paramorphic surrogate taking the form:

$$W = aX + bY + cZ$$

In reviewing the many studies using first-order model up to 1970, Slovic and Lictenstein (1970) state:

> In all of these situations the linear[1] model has done a fairly good job of predicting judgments, as indicated by r_j^2 values in the .80's and .90's for the artificial tasks and the .79's for the more complex real-world situations. (p. 36)

On the pro-configural side of the issue, continued efforts have been made to find situations where the higher-order and configural models will consistently out-perform the approximation afforded by the linear surrogate, superior performance being indicated by the amount of variance explainable by the policy model. In one study, Wiggins and Hoffman (1968) compared the performance of a quadratic model and a configural "sign model" to the first-order models for 29 psychologists in making a diagnosis from mental symptoms of 861 cases of MMPI (Minnesota Multi-Phasic Inventory) data.[2] The results showed that for 16 of the 29 judges, configural models were marginally better, but not to a statistically significant extent. In general, other such efforts have not been impressively successful in increasing the amount of explained variance and have caused some frustration to the researchers. This frustration has led the researchers to search for different statistical methods and devise new tactics in applying old tools. Among the methods attempted have been the application of Analysis-of-Variance procedures (Hoffman, Slovic and Rorer, 1968) (Slovic, 1969) and the "*a priori* partialling out" of variance due to the hypothesized configural effects (Valenzi, 1970). Even with these techniques, statistically significant evidence of configurality could not be established by the researchers,

---

[1] Much of the current literature is rife with the misuse of the term "linear" to denote "first-order" models as opposed to including all models with linear parameters.

[2] This set of data has become known as the "Meehl data" and has been used extensively by various researchers.

lending credence to Green's (1968) critique of all such efforts:

> Another difficulty with standard configural techniques analysis-of-variance, and the new index[3] is that they essentially are fishing expeditions. The experimenter will covet any configural effect, any interaction term he can find. He cannot begin to examine all the possible non-linear effects, and is very likely to miss those that are present, unless he knows where to fish. (p. 94)

Undaunted, other researchers have taken the approach of formulating completely new types of models and working with different decision situations in order to demonstrate configurality. The most prominent of these efforts has been the work by Einhorn (1970, 1971) in the formulation of the "conjunctive" and "disjunctive" models. Einhorn, hypothesized that the appropriate decision model is really a function of the type of decision task and other factors of the decision situation including: inter-judge differences, distribution properties of the cue values, number of stimulus objects, the correlation among the cues, the number of stimulus or cue dimensions to be evaluated for each object, and the context within which the decision is made. He stated that for certain types of situations, such as those in which the strongest attribute or the weakest attribute of the stimulus object was important, non-compensatory models of either the conjunctive or disjunctive type would apply respectively. The proposed conjunctive model is of the form:

$$Y = \prod_{i=1}^{N} x_i^{a_i} \qquad \text{or} \qquad \log Y = \sum_{i=1}^{N} a_i \log x_i$$

[3] The index referred to here is an index of cue consistency proposed by Hoffman (1968) as the "believability index" and relates to the degree to which separate cues match the preconceived ideas of how the separate cues should be correlated with the criterion.

and the disjunctive model is of the form:

$$Y = \prod_{i=1}^{N} \left(\frac{1}{a_i - x_i}\right)^{b_i} \quad \text{or} \quad \log Y = \sum_{i=1}^{N} -b_i \log(a_i - x_i)$$

In his experiments, Einhorn pitted these models against the first-order models for four judges in the task of ranking graduate school applicants based on Graduate Record Exam scores. He found his models gave significantly better results for three of the four judges and the shrinkage of his models upon cross-validation was much less than that achieved with first-order models for all four of the judges. This last finding is of special interest, in that Ward (1954) found that configural models generally did not cross validate as well as first-order models due to "over-fitting". Hence, high cross validities for a configural model would appear to enhance its claim to authenticity.

Einhorn's work has offered new evidence for the long fostered opinion that configurality existed, if it could just be uncovered. It is important to note that Einhorn himself recognized that his models would probably not perform as well in different decision situations. This admission did not placate Goldberg, however, who was apparently disgruntled by Einhorn's refutation of his previous assertion that first-order models fit best because they are mathematically simple and man thinks in simple ways. (Goldberg, 1968) Instead, Goldberg (1971) immediately set about the task of disproving Einhorn's findings on the much analyzed "Meehl data". The results of this vendetta showed that, as was nearly pre-ordained, the linear model did a better job than either the conjunctive or disjunctive models in depicting the judgment of clinical psychologist in making diagnoses. This merely tends to confirm Einhorn's hypothesis that the best model is a function of the decision situation. More importantly, in justifying his results,

Goldberg gave the first indication that he and other researchers, within the heretofore "pure policy capturing" camp, were awakening to the necessity for more emphasis on "representative" research designs as espoused by Brunswik (1956) would be necessary.

Bootstrapping:

Independent of the researchers belief or disbelief in HOW? the policy of a decision maker can best be modeled, there is always the problem of justifying WHY? they should attempt to model the decision maker at all. This debate has continued since Meehl's original proposal of the "actuarial combination of cues", with some clinicians, apparently justifying their jobs, claiming that the actual combination of the cues is only a small part of the overall judgment process. (Holt, 1970)

Most of the arguments for using policy models of the judge, as opposed to the judge himself, center around the consistency of a known and fixed computational formula as compared to the inherent variability of the human judgment process. One of the main proponents of using Policy Capturing has been Goldberg (1970) who described the human reliability problem by saying:

> He 'has his days': Boredom, fatigue, illness, situational and interpersonal distractions all plague him, with the result that his repeated judgments of the exact same stimulus configuration are not identical. He is subject to all these human frailties which lower the reliability of his judgments below unity. And, if the judge's reliability is less than unity, there must be error in his judgments--error which can serve no other purpose than to attenuate his accuracy. If we could...eliminate the random error in his judgments, we should thereby increase the validity of the resulting predictions. (p. 423)

In an analysis of the performance of first-order policy models versus the actual judgments on the "Meehl data", Goldberg (1970) found that the models were marginally better for all 29 judges than the judges themselves.

Dawes (1971) also demonstrated the superiority of policy models over actual judgments in predicting the rank of graduate admissions. Dawes and his fellow researchers at ORI have termed the phenomena of the policy model's superiority to the human in predicting judgments as "bootstrapping." The essence of the justification for "bootstrapping" is that mathematical models, by their very nature, are abstractions of the processes that they model. Therefore, if the decision maker's behavior involves following valid principles, but he is unable to follow them reliably, these valid principles will be abstracted by the model, and once abstracted, the superior reliability of the machine can be enlisted to follow them exactly, resulting in better quality judgments. It is recognized by this argument that this reduction in variability will occur if there is no systematic deviation from the principles based on the variables themselves, i.e., if the structural form of the model is correct. On this point, Dawes asserts that bootstrapping can occur even if there are structural defects of the model, such as in cases where a first-order model is used to represent a truly configural decision situation, if the cues themselves are monotonic and the data is fallible.

Dawes and Diller (1970) derived, and experimentally verified, a mathematical criterion based on the lens model parameters for determining when bootstrapping could be expected to occur for the case of first-order models of very good (near-optimal) judges. This condition for bootstrapping is:

$$r_j \sqrt{\frac{1-r_j}{1+r_j}} > \sqrt{r_{N_{j1}N_{j2}}} \sqrt{r_{Y_{j1}Y_{j2}} - \frac{r_j^2}{r_{\hat{Y}_{j1}\hat{Y}_{j2}}}}$$

where:

r_j = the predictability of the judge ($r_{Y_j \cdot \hat{Y}_j}$)

$r_{N_{j1}N_{j2}}$ = reliability of the residuals ($Y_j - \hat{Y}_j$), i.e., the correlation of residuals in cross-validation samples

$r_{Y_{j1}Y_{j2}}$ = reliability of criterion values in cross-validation samples

$r_{\hat{Y}_{j1}\hat{Y}_{j2}}$ = reliability of predicted correlation values ($\hat{Y}_j$) in cross-validation samples.

In this same study, they proposed amalgamating the first-order prediction of the judge from the policy model with the judge's actual prediction to get a hybrid prediction. This amalgamation is based on weighting the current judgment and the policy-model predicted judgment in the form:

$$A = \hat{r}_m z_{\hat{Y}_j} + \hat{r}_n z_{N_j}$$

where:

$z_{\hat{Y}_j}$ = the standardized "z-score" form of the predicted value $\hat{Y}_j$

z_{N_j} = the standardized "z-score" form of the residual $(Y_j - \hat{Y}_j)$

$\hat{r}_m$ = bounded validity coefficient for the model = $r_j / \sqrt{r_{Y_{j1} Y_{j2}}}$

$\hat{r}_N$ = bounded validity = $\sqrt{r_{N_{j1} N_{j2}}} \sqrt{1 - \dfrac{r_j^2}{r_{Y_{j1} Y_{j2}} \, r_{\hat{Y}_{j1} \hat{Y}_{j2}}}}$

Their experimental work on this procedure indicated that the amalgamation of predicted and judge criterion values was superior to the judge alone or the prediction alone, especially in the case where non-linear[4] cue use was considerable.

Bowman (1963) and Kunreuther (1969), have also proposed and studied the use of policy models of expert managers to reduce variance in management decisions relative to production management and inventory control. They have suggested that such a technique would be an excellent surrogate for the more traditional analysis, much of which is dependent on relationships of decision and responses that are greatly separated in time, or which is based on cost parameters that are intangible and must be estimated or assumed. This suggestion is tantamount to stating that it is more efficient to model the expert's opinion of the proper policy to follow, than to model the actual response as a function of the decision policy. This rather intriguing thought of using experts to estimate the cue-response relationships when the actual relationships were unattainable was also studied by Mullins and Usdin (1970). In that study, the use of judges to estimate the relation-

---

[4]Here "non-linear" refers to "higher order and configural" per the definition in Chapter I.

ships between the actual criterion values and the cues produced estimated criterion values that were virtually no different than the actual criterion values.

Sawyer (1966) analyzed the results of some 45 studies accomplished prior to 1966 and proposed that the judgment task be divided into two subtasks, measurement and prediction. He concluded that the perceptive capabilities of the human were best put to use in the measurement task and that the inherent reliability of the machine made the policy models superior in the prediction task. Sawyer's lead in decomposing the global judgment task into these two subtasks has more recently been followed by Einhorn (1972). Einhorn has agreed with Sawyer and proposes that not only should the measurement task be relegated to the expert judge, but that the whole process might be improved even further by using multiple judges in the measurement task with each judge being an expert in the measurement of some particular attribute. Such a technique would then use the expert measurements as an input to an actuarial combination of the cues in attaining a global judgment.

Apart from the justification of the relative accuracy and efficiency of using policy models, the WHY? of Policy Capturing can be answered relative to management's effective use of computers in the current man-machine era. Many authors such as Yntema and Torgerson (1961) have discussed management's need to "teach the computer how they want decisions to be made", and others state that in the operation of large organizations in the future, "specialist's knowledge will be stored in computer banks". (Vandell, 1970) Since computers are best at manipulating mathematical expressions, the use of captured policy models to instill man's value judgments in the computer appears to be the only logical approach to the task.

The Nature of the Decision Process:

In addition to studying the more mundane aspects of policy modeling such as the practicality of using these models as decision aids, a very considerable effort has been undertaken by psychologists in using Policy Capturing as a descriptive tool for understanding the fundamental cognitive processes of the decision maker. Some of these results have been the direct goal of specific research, while others have come about as by-products of the more applied research efforts. In any event, it is well to be cognizant of a number of these findings in any project in which a viable set of policy models is to be sought for use in a practical environment.

One of the original postulates in the area of cognitive processes was the concept of "cognitive strain" advanced by Bruner, Goodnow, and Austin. (1956) They hypothesized:

> Where the nature of a task imposes a high degree of strain on the memory and inference, the strategy used for coping with the task will tend to be less conducive to cognitive strain. To put it in terms of an analogy, if someone has to move a heavy weight, there is...likelihood that the mover will have recourse to strain reducing techniques for carrying out his task. (p. 112)

Another of the fundamental postulates resulted from the work of Miller (1956) who found that there are rather low limits on the number of items of data that the human can evaluate and integrate in a consistent manner.

In line with these postulates, Einhorn (1971) found that the judge made more consistent judgment when fewer cues were available. He proposed that man uses excess data in some hybrid fashion instead

of discarding it as might be expected in any effort to reduce cognitive strain. Similarly, Slovic and Lictenstein (1970) concluded that the results of past research showed that more information gave the judge a false confidence without improving his prediction consistency.

Slovic (1969) investigated the relative uses of first-order and higher-order cues in individual and group judgments. He found some evidence that when multiple cues were present, the tendency to use curvilinear models was less in the case of group judgments, while it was more in the case of individual judgments. This suggests that the averaging effect of group judgments may tend to obscure configurality in individual policies, especially if the configurality is different for different judges.

Earle (1970) studied the learning and interpersonal learning characteristics of judges trained on first-order and higher-order cues. He found that those judges trained initially to use a higher order cue could readily adapt to using a first-order cue without the aid of a teacher, but that those judges who had been trained on a first-order cue had to receive help in adapting to the use of a higher-order cue.

Another point of investigation has been in the area of sequential decision processes. Although most of this work has been carried out within the previously mentioned Bayesian Approach, Slovic (1966) and Hoffman (1968) have proposed that the judge first assesses the "believeability" of the cues and then proceeds to adjust the cue weights of his first-order policy model accordingly. In a more exhaustive work, Bettman (1969) essentially pursues this same notion except that he hypothesizes that the decision maker first assesses the consistency of the cues and given they are consistent, uses a simple or first-order model; however, if they are inconsistent, the decision maker uses a configural model in what Bettman calls a "problem

solving" approach. These findings are suggestive of the "lexicographic" model mentioned by Einhorn (1970) in which it is hypothesized that a decision is made on the most important variable first and then remaining variables are considered in the order of their importance. Unfortunately, Einhorn and others defer from pursuing this model due to their inability to mathematically model the process.

Summary:

Probably the strongest general impression to be gained from a review of the past research efforts in Policy Capturing is the artificial nature of the environments in which the research has generally been conducted. This is especially paradoxical in that one of the basic tenents of the Brunswikian Lens conceptualization is the use of "representative" designs. Only recently, researchers such as Goldberg (1970), Einhorn (1972), and Dawes (1971) have begun to voice concern over this departure from reality, with Dawes expressing his reservations of past research results in the statement:

> Most of these studies have been conducted in analogue experimental situations and it is reasonable to ask whether the situations were constructed in such a way that the linear model works, but are not representative of the real world decision making situations. (p. 181)

Most of the past analyses have been performed on decision tasks where only a limited number of well-defined cues have been used. As a typical example, Valenzi (1970) arbitrarily reduced the number of cues in his study from 20 to 5 based on a preliminary interview with one subject in the study. Such practices would seem to be risky in the face of data indicating that even experienced judges have a difficult

time accurately describing the relative weights of the variables in their policies (Slovic and Lictenstein, 1970). Valenzi's own results indicated that only one of the four judges in his study was able to correctly estimate the relative importance on the variables in their policies.

In most studies even the procedure of asking the subject which variables they "felt" were important, before deciding on a predictor set, has been omitted with the subjects being given a set of predefined cues. This practice, which is equivalent to ignoring steps 1 and 2 of the Policy Capturing process as depicted in Figure 1-1, page 7. was first challenged by Garner (1970) when he noted that the psychologists modeling clinical judgments had paid virtually no attention to the problems of identifying, scaling, and coding appropriate stimuli. Einhorn (1971) later asserted that:

> The practice of presenting cues to the judge in a decomposed form not only imposes the experimenter's own judgment as to what the relevant cues actually are but more importantly it does a considerable part of the cognitive work <u>for</u> the judge. (p. 8)

A related problem with the artificial environment is the extent to which the experimenters have relied upon the statistical and computational niceties of factorial designs in choosing their stimulus data sets. In an early study of the possible effects of using sample data sets that were statistically different than the underlying population, Dudycha and Naylor (1966) perturbed the predictor intercorrelation matrix from the true values and studied the resultant effects on the policies derived for the judges. They found that perturbation of the interrelationships between predictor variables to minimize the correlations resulted in lower rater consistency and better discrimination of

the effects of the individual variables, while increasing these correlations resulted in higher rater consister :y and worse discrimination at the individual variable level. This would indicate that artificial orthogonalization of the predictor variables by use of factorial designs with equal-cell frequencies, would tend to produce policies that were less predictive than the; should be. Hence, computational efficiency has been purchased with decreases in accuracy in such experiments.

In addition to these rather arbitrary decisions relative to which variables and which data units to use in the experiment, the criterion for evaluation of the various policy models has usually been restricted to using the percentage of explained variance, as expressed by the squared multiple correlation coefficient (r_j^2). Slovic and Lictenstein (1970) found that most of the models that they reviewed were <u>NOT</u> cross-validated. Further review by this investigator, indicates that only in the case of those models developed or sponsored by the researchers at the AFHRL, and in a few other cases such as Bettman (1969), has cross-validation of the derived policy equations been accomplished on a systemmatic basis. Admittedly, explained variance is a commonly used "measure of goodness", but with small data samples models can achieve inflated values of r_j^2 solely due to the idiosyncratic nature of the sample data.

A second general impression of the past work is the rather restricted range of modeling possibilities that have been entertained. Within the correlational paradigm and the Brunswikian lens framework, almost all of the work has been confined to linear first-order, higher-order, and configural models of the type:

$$Y = \sum_{ij} b_i x_i + \ldots + b_j x_j + b_{ij} x_i x_j + E$$

where the values of b_i, b_j, and b_{ij} are considered to apply over the total predictor space, V^n. Hoffman (1968) characterized such models as being n-dimensional hyperplanes and states:

> The concept of a best-fitting hyperplane implies sort of an averaging of the effects of all points in the multiple dimensioned space, and, since linearity is assumed, the regression coefficients are invariant over the entire stimulus space. (p. 59)

Even Einhorn's (1970) departure from the more standard main-effects or main-effects-plus-configural-effects modeling approaches by using the conjunctive and disjunctive models still represents the fitting the entire predictor space by a single, continuous hyperplane.

The only works that reflect a departure from this concept are the efforts of Slovic (1968), Hoffman (1968) and especially Bettman (1969). In the first two of these studies, it is proposed that the judge adjusts the coefficients in his first-order model based on an initial assessment of cue-consistency or "believability", while in the latter study, the whole structure of the policy model is dependent upon this initial assessment. These procedures effectively divide the predictor space into a set of discontinuous hyperplanes, each hyperplane representing a local policy or a local model. Bettman (1969) appears to be the only one of these researchers to recognize this implication with his conclusion:

> The major and most useful conclusion to draw from the above ...(research)... is that a linear or less configural model can be viewed as a complex model conditioned by a given cue configuration. (p. 134)

The real potential of the concept of modeling the entire predictor space, V^n, with a series of disjoint hyperplanes over subspaces, S^m, of the predictors has not been further addressed. However, this technique presents a possible solution to the problem of obtaining models that retain the desirable characteristics of first-order models while accounting for the configural logic processes that many judges feel they use. One possible reason that this area has remained unexplored is the problem of determining the basis upon which to deaggregate the total predictor space into appropriate subspaces. Such a division would surely have to be based on the inherent characteristics of the policy and the predictors for it to be meaningful. Any arbitrary approach would be analogous to Green's (1968) "fishing expeditions" and would be doomed to frustration. Blaylock (1969) voices a similar opinion in his discussion of building theories with multivariate models by stating:

> One cannot simply build theories involving interaction effects without some rationale to suggest the conditions under which non-additive relationships might be expected. (p. 156)

Thus, a viable methodology for identifying appropriate subspaces over which local models could be fit would appear to be a necessary precursor to the exploration of the approach of using multiple hyperplanes, as opposed to using a single hyperplane, in the modeling of policies.

As a final comment in this review, the observation of Slovic and Lictenstein (1970) that resulted from their own exhaustive review of the previous research is germain. They concluded that for as much work as had been accomplished in many different areas of

judgment modeling, surprisingly little cross-communication among the efforts had transpired. They compared this situation to a speeding vehicle without side-windows, the only visibility being directly down the path of travel of a given paradigm, or modeling strategy. With this comment they have noted the requirement for a side-window; the review of the recent work of others such as Hoffman, Bettman, and Einhorn suggests a faint outline as to where a side-window might be carved; it is the purpose of the work which follows to open this side-window so that future researchers can look out and the practitioners of Operations Research can look in.

CHAPTER III: FOCUS OF THIS RESEARCH

Application of Policy Capturing in Practical Environments:

The introduction of the Policy Capturing Concept and a review of the past work can provide the genesis for many tantalizing ideas of the potential benefits that such a technique could offer to Managers, Educators and other Administrators in extra-laboratory environments. However, as with any scientific concept, a problem of the conversion from a fairly general collection of thoughts to a specific methodology exists in the adaptation of the Policy Capturing Concept.

The review of past research pointed out that such a conversion has only been previously attempted in a few instances, specifically, those studies undertaken by the researchers associated with the Air Force Human Resources Laboratory (AFHRL). In these cases the effective implementation of the concept has been materially aided by the particular decision tasks and the environments involved; namely, decision tasks in which the predictors were such quantities as test scores, aptitude indices, and performance ratings; data sets that consisted of large quantities of data; underlying judgment processes that were very well represented by first-order policy models; and, minimal emphasis upon correlation of the paramorphic models with the actual logical procedures of particular judges.[1] Thus, even the AFHRL work has been

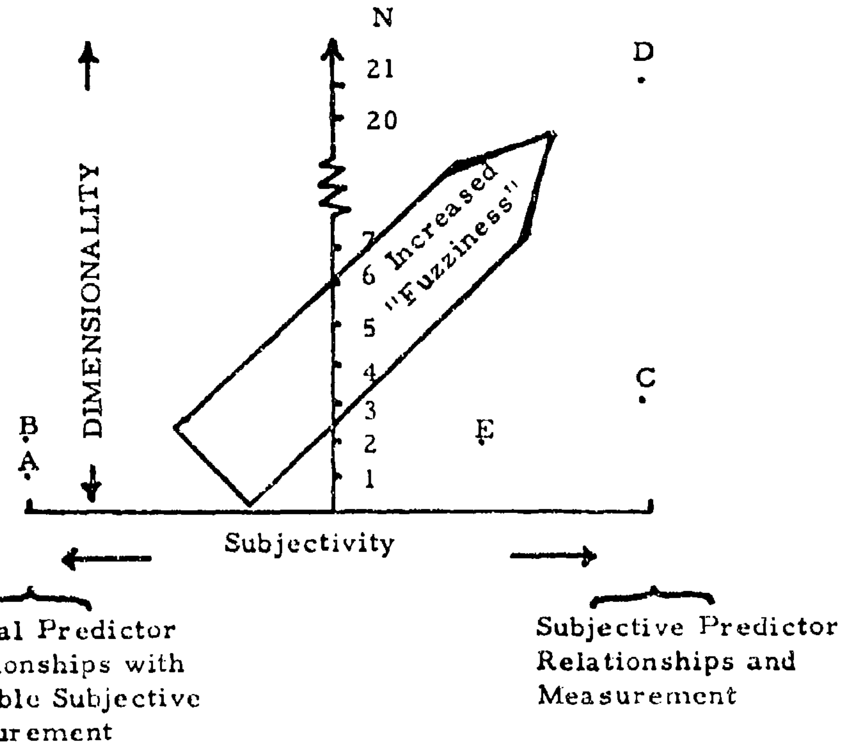
[1] In the latter aspect, the difference between a policy model that is visible and can be explained and a policy model that is reflective of a particular judge's self-perceived logical process is important. The only requirement for the former is that the variables and structural relationships in the model be relatable in some rational manner to the decision process; the requirement in the latter is that those relationships must not only be rational, but also, they must be in harmony with the way the individual judge "believes" he makes decisions.

in environments with attributes that make them pseudo-clinical in nature since they are closer to ideal experimental situations than can be generally expected in the business, academic, and governmental or other administrative types of environments. In the extra-laboratory environments, categorical predictors, limited data sets, complex and configural logic processes, and skepticism for "black box" answers are probably more characteristic of the decision situations.

The focus of this dissertation is directed at exploring the hypothesis that Policy Capturing can be applied effectively in the extra-laboratory environment. The fundamental assumption underlying this hypothesis is that the successful application of Policy Capturing to practical decision tasks is dependent upon providing the decision modeler with an efficient methodology for deriving explainable policy models that reflect the mental procedures that the decision maker feels he goes through in making judgments. This assumption is the basis for the first part of the effort of the research, i.e., the analysis of various modeling alternatives and the development of a methodology that will lead to models that are both accurate and descriptive of the underlying logical processes. The stated hypothesis is the basis for the remainder of this research, i.e., the application of Policy Capturing in an extra-laboratory environment and the analysis and demonstration of the benefits attainable by applying the Policy Capturing technique in this environment.

At this point, a further characterization of which types of decision tasks would probably be most amenable to the use of Policy Capturing is warranted. In making such a characterization, one might conceptualize all decision tasks as lying in a two-dimensional envelope such as is shown in Figure 3-1. One axis of this envelope corresponds to the inherent subjectivity of the decision situation and the other axis

CHARACTERIZATION OF DECISION TASKS



EXAMPLE JUDGMENTS

A, B: Computation of Perimeter Length; only judgment involved is in reading measurement scale, A--perimeter of a circle; B--perimeter of a rectangle

C: Acceptability for Graduate School (Dawes, 1971); two judgments involve weights for variables and evaluation of the quality of the undergraduate school (QS).

D: Probability of Being Hired as Secretary (Valenzi, 1970); up to 20 different predictors could have been used.

E: Evaluation of Job Difficulty (Mead, 1970); both subjective and quantitative variables used to predict job difficulty.

FIGURE 3-1

corresponds to the dimensionality of the predictor space. Any decision situation in which the relationship between the predictors and the criterion cannot be fully explained by some immutable natural law would fall somewhere within this envelope. Any decision situation within the envelope has a requirement for some degree of human judgment in making the final decision. At one end of the subjectivity axis, the selection of which equation or physical law should be applied is the only part of the total decision left to the judge, while on the other end of this axis, the judge is totally responsible for the subjective identification of the relevant variables, the assessment of their individual relationships with the criterion, and the combination of their contributions into a global judgment.

Policy Capturing would be particularly applicable in the region of this envelope where considerable fuzziness exists in the relationship between predictors and the criterion, and in that region where the dimensionality of the problem is high.

The more mechanical aspects of decision scenarios that solicit the application of the technique are:

1) The decision task is routine and repeated on a frequent basis with the relative importance of the various predictors remaining stable in the short term.
2) There are multiple decision makers for whom a composite decision is desired or whose individual decisions should be consistent.
3) The quantity of the data for each decision task might lead the judge to simplify the decision process and focus on different subsets of the cues from case to case in an effort to reduce cognitive strain.

4) The decision process is possibly subject to biasedness and the detection and removal of this biasedness would be beneficial.

5) The actual predictor/criterion relationship can only be determined after a considerable time lapse and such a time interval can render any model of this relationship inappropriate and obsolete due to environmental changes.

6) The relative importance of the parameters of the decision process are not known and are not estimable from explicit physical relationships.

Given these basic characteristics of "fuzzy" decision processes and the mechanical aspects of target applications, a more detailed discussion of several functions that Policy Capturing serve is possible. Five generic roles in which Policy Capturing would be beneficial are:

A. <u>Factoring Expert Judgment into Automated Decision Making</u>

As noted in the Introduction, the Operations Research practitioner has long been concerned with optimization of value based on models of various operational systems, but these models often include imprecise "guess-timates" of Management's utilities for risk and profit in the form of objective function coefficients. The more precise approach of implicitly determining a manager's expert opinion of the relevant variables and their weights via Policy Capturing would assure that the essence of the manager's expertise is factored into the decision model and not just those aspects of his experience that are easy to quantify or verbalize. Whereas the objective functions of many organi-

zations are really a reflection of the composite judgment of managers at many levels, the use of explicit policy models for the contributors would provide an improved basis for discussion and integration of policy opinions of each manager in a manner commensurate with his proper role in the company. With less explicit procedures, a person's impact upon organizational policies are often based solely upon his debating ability, dominant personality or title, and not upon the quality of his opinions.

B. Standardization of Policy

The use of Policy Capturing techniques to assess intra-judge consistency, inter-judge consistency, and even inter-organizational consistency would appear to have merit. On an intra-judge basis, the inability of a given judge to make consistent judgments over a set of similar judgmental situations could indicate a fundamental uncertainty in his perception of the task or possibly a higher susceptibility to random external pressures. On an inter-judge basis, the assessment of the consistency of two or more persons performing identical decision tasks would help to make the performance of this task invariant to which judge makes the decision.

The use of Policy Capturing to detect systematic differences on an inter-organizational basis is particularly intriguing from the standpoint of administering a large de-centralized organization. In cases where a policy is set at some central headquarters, but administered at various subordinates levels such as divisions of large corporations or regional offices of government agencies, Policy Capturing could provide a vehicle for identifying communications and interpretation problems between the policy makers and the policy administrators.

C. Training

The use of Management's policy models would also be of value in the training of new personnel. This is especially true where the judgment process is highly subjective and Management is continually faced with the dilemma of whether to use the experts to "do-the-job" or to "train-the-new-people". Codifying the expertise of the seasoned employee in the form of policy models and using these models as training aids would enhance the transfer of knowledge and experience without requiring continual participation of the expert judge. Such models could also be used as a baseline or standard with which experienced employees, who may have lost contact with the lower level decision making process, could "recalibrate" themselves.

D. Policy Simulation

Once a policy has been captured, the modification of the policy model to simulate changes in the policy could provide a valuable tool for Management in performing a "sensitivity analysis" and assessing impacts of such changes on an *a priori* basis. This application was proposed by Ward and Davis (1963) as an efficient method for analyzing many possible policy alternatives before institution of any changes, especially in cases where the effect of a change might take a considerable time to become evident.

E. Mechanization of Routine Judgments

The phenomena of "bootstrapping" and the inherent reliability of machines presents possibly the greatest opportunity for benefit to be gained from Policy Capturing. Not only does the reliability of the machine, in evaluating and integrating numerous stimulus variables, offer advantages, but further, the potential increases in speed and efficiency to be gained in performing the function with minimal human

intervention are substantial. The use of the machine allows for the analysis of more data than can be used by the human on a continual basis. For example, in the only previously published estimate of savings to be gained by using policy models, Dawes (1971) estimates that approximates $18 million could be saved annually, on a national basis, by mechanizing the evaluation of graduate student applications using captured policies of admission boards. Substantial savings can be imagined even if the computer does not make the final decision, but only pre-processes the data and provides its evaluation to the judge in such a form that allows the judge to make his decision more efficiently.

The Transition from Hypothetical to Methodological:

Speculating and hypothesizing as to the potential of a concept can generate considerable enthusiasm, but such activities do not constitute verification that the concept is anything other than a dream. If a concept is purely theoretical, the researcher can set about defining ppstulates, lemmas, and theorems which will result in necessary and sufficient logical proofs of the concept's validity. However, if one is dealing with such "fuzzy" entities as man and the human judgmental process, he is confined to the world of empiricism wherein the only available approach to verification is demonstration. Given numerous repeated demonstrations over a broad range of conditions, these empirical demonstrations can result in acceptance and tacit verification of the concept. However, such verifications only hold within the bounds of the underlying assumptions and population characteristics for which the demonstrations were performed, a point often overlooked when assumptions become so commonly used that they cease to be emphasized in the analysis and description of the results.

Many researchers who find themselves operating in the world of empiricism try to regain some measure of "theoretical legitimacy", or at least try to simplify the verification procedure, by relying upon the theoretical developments of mathematical statistics and demonstrating "statistical significance". This too, requires the making of assumptions which often go unsubstantiated, being justified only by the claim that they are commonly used. Among the most commonly used assumptions are those of randomness, univariate or multivariate normality, independence, and homoscedacity. The carte blanche use of these assumptions is one of the major problem areas in translating many of the findings from the clinical environment to operational environments. If one is to effectively apply his results to the real-world he must derive these results under real-world experimental conditions or demonstrate that any underlying assumptions do not vitiate his results in the real world.

Criteria for Attaining Effective Models:

Consistent with the philosophy of performing Policy Capturing with ecologically valid data and being fully cognizant of any assumptions, the criterion for the measurement of the quality of a model must also be congruent with the requirements in the environment being modeled. Establishment of such requirements represents a further delineation of the previously stated fundamental assumption of this research. The basic characteristics that models should have to be viable in the business, academic and administrative fields are:

A) Accuracy: The models must possess at least that degree of accuracy necessary to produce a "bootstrapping" effect and to provide an improvement in operational efficiency.

B) Reflectivity and Interpretability: The models should be reflective of the logical processes of the judge in order that he may identify with the model and have confidence in it.

C) Communicability: The model should be readily communicable if it is to be used as a training device. Simplicity in a model would add to its communicability to and by non-scientifically oriented judges.

D) Data Amenability: The model must be able to accept and use data that is defined and categorized similarly to the manner in which the judge perceives and uses it in his normal judgment procedure.

E) Data Conservatism: The modeling procedure should be operable with data quantities that are not prohibitively large in terms of data collection and should not require major disruption of the on-going decision process for data collection.

The above delineation of the characteristics of effective models helps to clarify the criteria for measuring model goodness, but the very nature of these characteristics does not readily admit them to quantitative measurement. The researcher is still left with subjectivity in the evaluation of various models; however, the dimensions of this subjective judgment have been better identified.

These criteria are not only qualitative, but their applicability is also dependent upon the functional application that is intended for the model. Table 3-1 reflects an attempt by this researcher to list the three most important model characteristics for each of the functions that were identified previously.

TABLE 3-1

FUNCTION	MODEL CHARACTERISTICS
Factoring "Expert Judgment"	A, B, C
Standardization of Policy	A, B, C
Training	B, C, D
Policy Simulation	A, D, E
Policy Mechanization	A, B, D

The Research Approach:

The foregoing definition of the experimental arena and the research objectives is a necessary prelude to the definition of the specific research approach and research tasks to be accomplished. The two major objectives of the effort are:

1) the analysis and development of a modeling methodology that will produce policy models with the desired characteristics, and,

2) the demonstration of the potential viability and utility of policy models of an actual decision process in an extra-laboratory environment.

This research effort is unique with respect to past judgment analysis studies since it includes the pursuit of both of these objectives simultaneously. The effort to adapt and apply a new methodology for defining policy models is actually embedded within the effort to obtain and validate policy models in an operational environment. In light of the current state-of-the-art in judgment modeling, the success of the methodology development effort is considered a pre-requisite for the accomplishment of the second of these objectives. In terms of the

functional steps of the traditional Policy Capturing process as depicted in Figure 1-1, page 7, the total research effort represents the tailoring of all of the functions to accommodate the conditions in an extra-laboratory decision situation. More importantly, the methodology effort involves the development of a completely new procedure for step #3 of that process.

In this respect, two separate contributions are expected from this research. The first is a set of procedures or guidelines for constructing policy model in extra-laboratory decision situations. These guidelines and the discussion of the model building methodology are addressed in detail in Chapter IV. The second contribution consists of the facts and implications that are generated in the process of conducting the Policy Capturing exercise in the particular decision situation. The background of the decision environment and the results of the Policy Capturing exercise are discussed in the format of a case study in Chapter V and Chapter VI.

An Overview of the Research Tasks:

The fundamental task of developing a model building technique is accomplished through the definition of "structural image" models with the Automatic Interaction Detection (AID) procedure. AID was recently developed by a group of sociologists at the University of Michigan for the purpose of discovering structural relationships between variables in large quantities of sociological data.

The motivation for pursuing this approach is the similarity between the conceptual diagram of a configural-effects judgment process, as shown in Figure 3-2, and the binary tree that results from the application of the AID technique. Figure 3-2 schematically diagrams the configural effects judgment process in which the relevant stimulus

A HYPOTHETICAL "CONFIGURAL-EFFECTS" JUDGMENTAL PROCESS

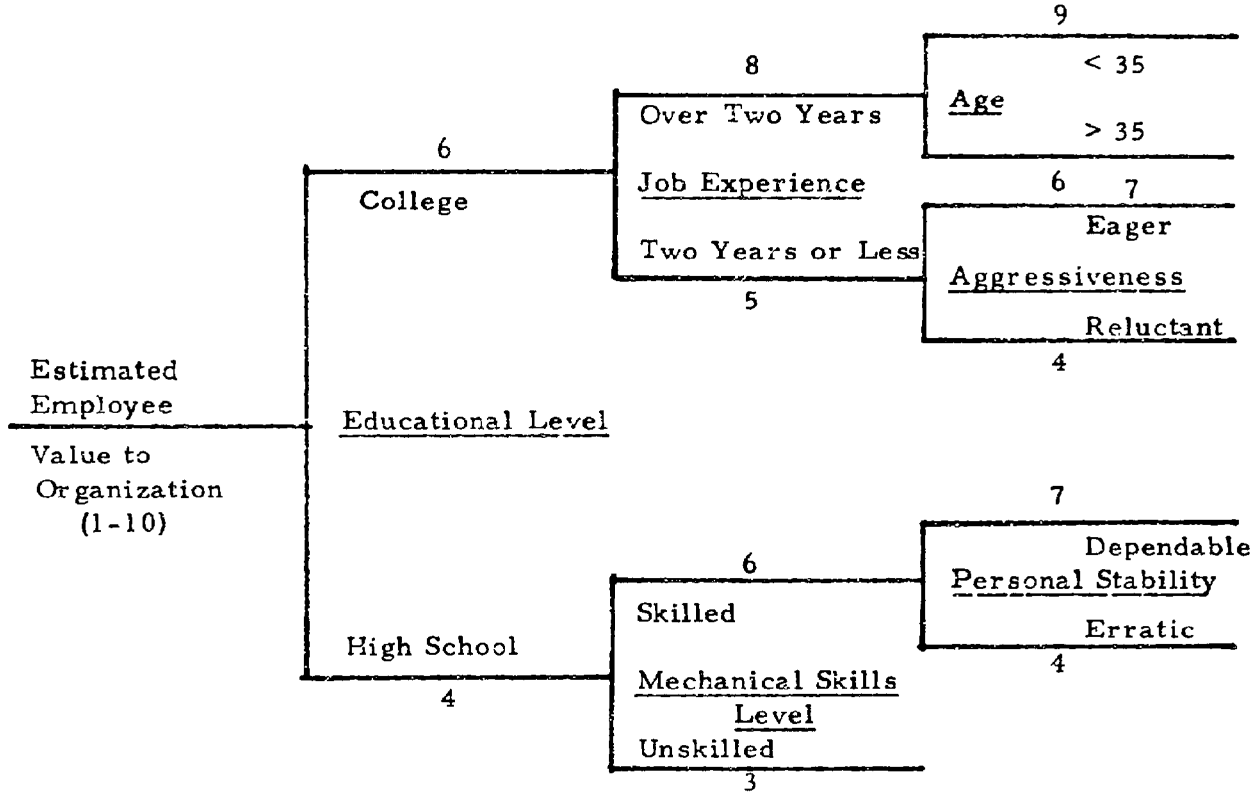


FIGURE 3-2

variables and their relative importance is obviously dependent upon the value of the previously considered stimulus variables. In a conceptually similar manner, AID performs a "local analysis" of the data and builds a binary tree that reflects the "structural image" of the data. In this tree each node represents a subgroup of the total data set. Each subgroup is uniquely defined by a common set of predictors and, each split or branching reflects the most important variable relative to reducing the unexplained variability in the data at that node. Hence, the splitting process reflects the importance of the predictor variables relative to those predictor variables defining the subgroup or node. If interaction and configural relationships exist among the predictors, it can be detected via analysis of the predictors involved in the splits and the relative importance of each predictor after each split.

In this research, a conceptual link between the "structural image" model of AID and configural judgments processes is made in terms of modeling judgment processes with a series of local hyperplanes. A local hyperplane is a hyperplane defined over a subspace of the predictor set.

The major subtasks within this effort that relate specifically to the building and application of policy models include:

1) Identification of an extra-laboratory decision situation in which most of the potential benefits from policy capturing could be derived.
2) Identification of <u>all</u> potentially pertinent variables based on discussion with actual decision makers.
3) Collection of data representative of the population of decision cases.
4) Coding, analysis, and refinement of data categorizations and predictor definitions.

5) Collection of decisions by each participating judge on codified predictors over the data set.

6) Investigation of alternative modeling approaches.

7) Development of models for each of the participants.

8) Verification and field-testing of models in an actual extra-laboratory environment.

9) Analysis of benefits and problems in actual application of the policy models.

The decision process chosen in subtask 1 as the basic vehicle for this research is that of the installment credit officer in granting or denying installment credit on durable appliances. This decision is based on credit applications forwarded to the credit institution by the appliance dealer and does not involve direct personal contact between the decision maker and the applicant. Several of the more pertinent reasons for using this situation as the basis for the research are:

1) The decision process is felt to be configural in nature.

2) The data and decisions are real and typical of the types of data and decisions made throughout business, academic, and governmental agencies.

3) The decision task is a routine and frequent part of the judges daily duties.

4) The judge has an economic motive for making the proper decision in the most efficient and accurate manner and is willing to use any decision aids that are understandable and effective.

5) Most of the hypothesized benefits of Policy Capturing could accrue in this environment.

6) Attainment of meaningful results in this application should constitute a demonstration of the general viability and utility of the Policy Capturing methodology in the extra-laboratory environment.

Subtasks 2, 3, 4, and 5 are accomplished with the objective of minimizing the influence that the procedures and decisions of the data definition, collection, or coding phases have upon the derived policy models. Particularly important aspects of these subtasks include:

1) Initial identification and coding of the predictors is accomplished through extensive discussions with the loan officers involved with refinement and recoding of the predictors variables resulting from the initial analytical efforts,

and,

2) The decision data consists of a random sample of approximately half of the actual decision cases that occurred during a five-month period from May to September, 1971. Actual decisions made prior to the start of the project are used in one part of the analysis. Decisions subsequently collected for each of five loan officers on the codified version of the predictor variables are also used.

This emphasis in collecting data and decisions "au naturel" represents a departure from the usual procedures of policy modeling studies which work with predefined predictors and fixed categorizations. This approach complicates the whole Policy Capturing procedure, but hopefully justifies itself by resulting in models that are more realistic and easier to implement in actual environments.

Subtasks 6 and 7 involve the development of first-order and second-order policy models for the judges and the comparison of these models with the "local models" attained through the application of the techniques resulting from the methodological development effort.

Finally, the collection and use of cross-validation data by the field test approach in subtask 8 is unique to this study. This subtask not only provided a measure of cross-validation of the derived models, but also, identified potential implementation problems and served as a demonstration of the viability of the concept to the participants.

Summary:

The focus of this research project is unique in that it not only involves the entire Policy Capturing process, and the application of the entire process in "real-world" environments, but it also involves a major methodology adaptation as an integral part of research.

The primary goal of the research is obtaining policy models that are "implementable". The criteria for "implementable" models are subjective at best, but a set of desirable model characteristics has been proposed in this chapter. The absence of these characteristics may not dictate inviability for a model, however, the fundamental assumption of this research is that possession of these attributes will enhance the viability of a model. As noted by Rosenthal (1966), such an assumption reflects a bias in the attitude of the investigator and could be a significant determinant in the results. Therefore, the research approach has been formulated so as to enhance the credibility of the results as much as possible by modeling actual decision makers who were making actual decisions in their natural environment, and then subjecting the resulting models to the most rigorous of all validation tests -- analysis of their performance under actual environmental conditions.

CHAPTER IV: THE DEVELOPMENT AND USE OF STRUCTURAL IMAGE MODELS IN POLICY CAPTURING

Introduction:

This chapter presents the conceptualization, an example, and guidelines for the modeling of configural judgment processes as a series of local models which define hyperplanes over subspaces of the predictor set. The adaptation and use of the AID algorithm as a tool in portraying the structural image of the judgment process and in defining appropriate subspaces and hypothesizing local models is discussed.

The technique is applied to define local models for the judgment data published by Valenzi (1970) as well as to define the policies of installment loan officers. The results of the former analysis are compared with the results published by Valenzi for his second-order regression and ANOVA models. The modeling effort for the loan officers comprises a major effort for this research and is discussed in detail in Chapters V and VI.

Conceptualization of the Configural Judgment Model as a Series of Subspace Hyperplanes:

Past modeling of judgment processes as either first-order, higher-order, or configural linear statistical models has generally been in the context of "continuous" models. In this discussion, the descriptor "continuous" has the following connotation:

1. If a model is of first-order, the coefficients of each of the predictors applies to all values which that predictor can assume.
2. If the model has interaction terms, the configural relationship between the interacting variables is assumed to be the same over all possible values which the individual variables can assume and the coefficient for the configural term is applicable to all possible combinations of these values. Hence the adjective "continuous" is relating to the domain over which the terms of the model are defined and does not relate solely to the "linearity" of the model coefficients.

Continuity has been inherently imposed through the formulation of the judgment models as linear regression models of the form:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_{12}x_1x_2 + \epsilon$$

where the coefficients, "b_i", apply over the entire predictors space, V^n.

As discussed in Chapter II, Hoffman (1968) and others have described such models as representing the "best fitting hyperplane" over the predictor set with the coefficients reflecting the "average" effects over the entire stimulus space.

The near-universal representation of judgment processes as "continuous" models appears to largely have been a function of the availability and the understanding of the standard regression computational routines and the types of predictors that were considered in the

models. The alternative of formulating judgment models in terms of discrete or binary regression models has been mysteriously avoided. This avoidance may be due to the fact that the predictors considered in past studies have generally been interval variables. In those cases where there were categorical predictors, the approach usually has been to attempt various multidimensional scaling techniques to transform the variables into interval form.

There are certain advantages to "continuous formulations," two in particular being the economy of degrees of freedom required by the predictor vectors, and the parsimony in describing models that have fewer predictors.

On the other hand, binary or categorical regression formulations, in which each possible value of each predictor represents a separate category, consumes one degree of freedom per category and results in one predictor vector and one coefficient for each category. Such models do not represent hyperplanes, but rather, a series of unique points in the stimulus space. The advantage of binary models is that they inherently avoid the assumption of continuity and each of the coefficients reflects the exact effect for a single point in the stimulus space. Thus, there is no "averaging" of the effects represented by a particular coefficient.

In the past research, the use of ANOVA models by the researchers at the Oregon Research Institute and by Valenzi (1970) are the only serious ventures into the use of categorical models.[1]

---

[1] Appendix A discusses equivalency of ANOVA and categorical/binary regression models.

Clearly, a compromise between the de facto assumption of continuity and the alternative of considering each point in the predictor space is desirable to reach an optimum tradeoff between parsimony and exactness.

A viable approach to such a compromise appears to lie in the definition of a mixed model in which both the parsimony of the continuous model and the exactness of the discrete model are utilized. In terms of hyperplanes and subspaces, this approach envisions defining a number of discrete subspaces, each of which can be modeled adequately with its own hyperplane.

In reference to judgment processes this corresponds to defining subprocesses over various patterns, or subsets, of the cue variables. Past researchers, Slovic (1969), Hoffman (1968), and Bettman (1969), actually provide the genesis for this type of conceptualization with their suggestion that judges may actually think in a sequential fashion. Their contention that man first surveys the predictor values, assesses their consistency or believability, and then selects and applies the appropriate decision logic is conceptually analogous to the initial definition of a particular subspace, followed by the application of a model within that subspace.

In terms of the "structural image" of a configural effects judgment process as shown in Figure 3-2, page 53, the subspaces could be represented by the branches of the tree and the hyperplanes are defined by those variables considered in any branch.

At the left of the tree, one subspace is defined as consisting of all possible categories of the individual predictors and all possible combinations thereof. After the first branching, two subspaces are defined on the basis of the predictor used in that stage. In terms of

Hoffman's (1968) work, that predictor may be some index of "believability"; in terms of Bettman's (1969) work, it could be an index of cue "consistency". Each sequential branching results in a further de-aggregation of the predictor space, and in the limit, a completely categorical model would be defined in which each subspace consisted of a single pattern of predictor values.

If one pursues this analogy between a sequential decision process and the division of the entire predictor space into subspaces, the major problem in modeling the process becomes one of defining the appropriate subspaces. This problem is combinatorial in nature. Considering a decision process with five trichotomous predictors, there are 3^5(243) individual predictor combinations, or single-point subspaces.[2] Furthermore, there are 2^{243}-1 (a very large number!) possible combinations of these points into unique patterns of subspaces. Obviously, a trial and error search would be prohibitive and an efficient method of defining the subspaces is required.

The Automatic Interaction Detection Algorithm:

The Automatic Interaction Detection (AID) algorithm was originally developed by J. A. Sonquist and J. N. Morgan of the Institute for Social Research, Ann Arbor, Michigan. In a series of monographs and presentations, they proposed, explained, tested, and expanded the computer code for the algorithm. (Sonquist and Morgan,

[2] This would be the case if enough data points existed such that each possible combination of the predictors existed in the data set. Otherwise, the maximum possible number of combinations is the upper limit on the number of single point subspaces.

1964)(Sonquist, 1967) (Sonquist, 1970)(Sonquist, Baker, and Morgan, 1971)

The algorithm accomplishes the sequential division of the data set into subsets based on that split which causes the greatest reduction in the unexplained variability. The procedure involves the iterative application of one-way analysis of variance over every possible split in the predictor set. The major logical steps of the algorithm are depicted in Figure 4-1; a detailed mathematical description is given in Appendix C.

The output from this computer program consists of a series of splits and a set of terminal subgroups which best explain the variability in the criterion variable. A model of the form:

$$Y = \sum_i b_i x_i + \epsilon$$

where: b_i = the mean response of those data units falling in subgroup "i".

x_i = a binary group membership flag for those data units falling in subgroup "i".

can be written for these subgroups.[3] A major benefit of the program is the graphical display of the splitting process in the form of an AID-Tree as shown in Figure 4-2.

This graphic display of the splits was one of the primary motivations for developing the program since it allows visualization

---

[3] The relationship between the AID model and regression models are discussed in Appendix A.

LOGIC OF AID ALGORITHM

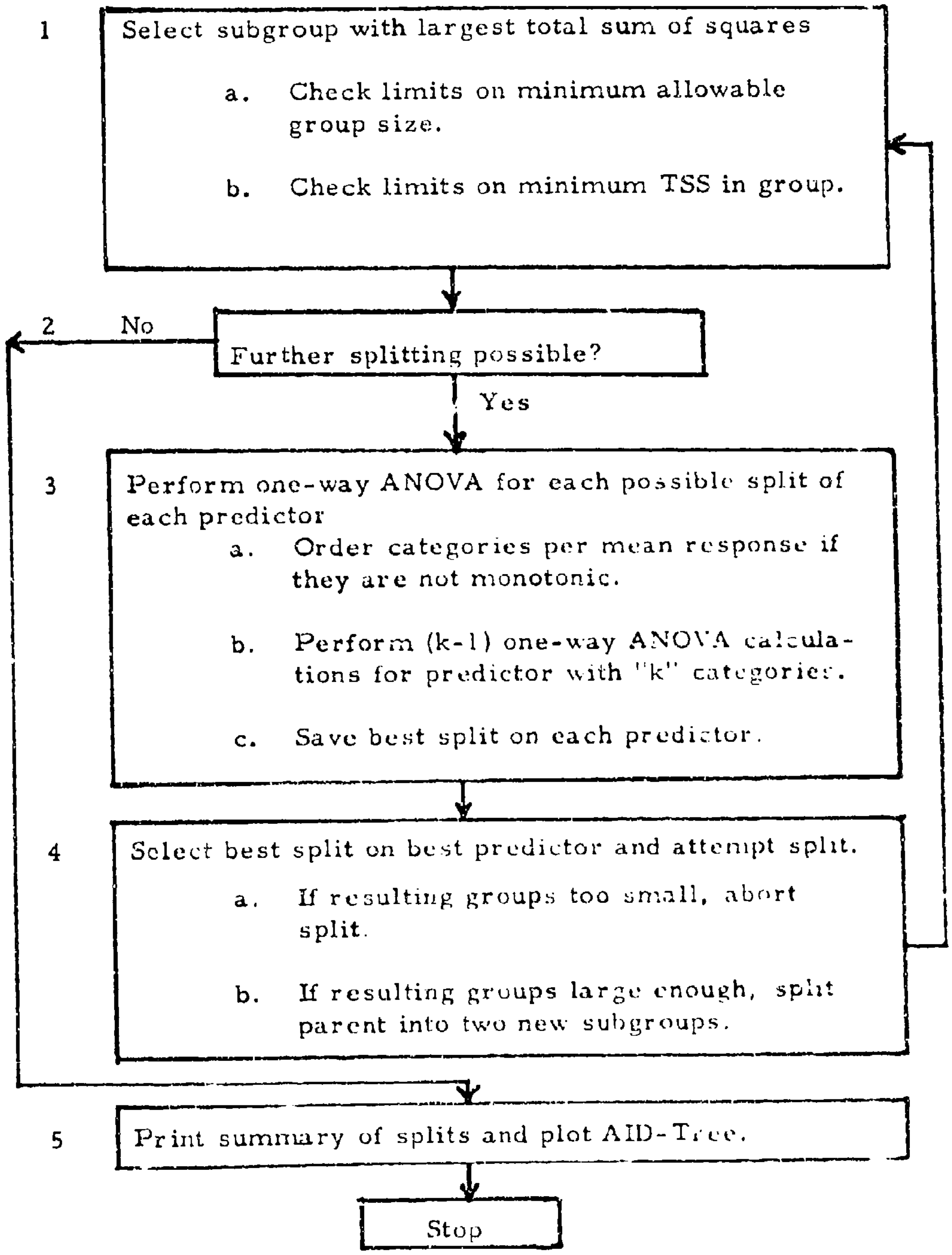


FIGURE 4-1

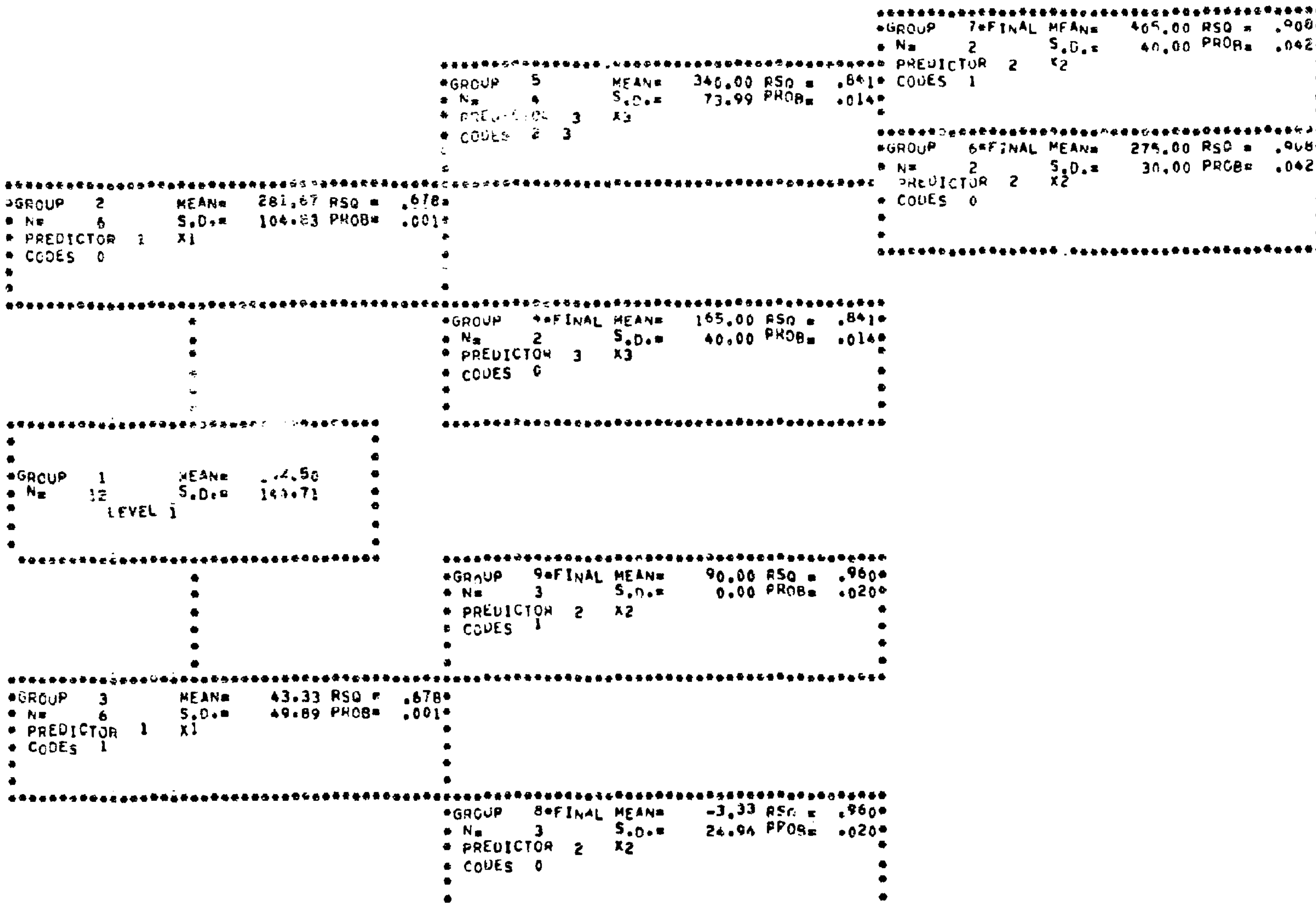


FIGURE 4-2, EXAMPLE OF AN AID-TREE

of the inherent structure in the data. A second motivation was the elimination of the requirement for assuming that linearity and additivity were properties possessed by the proper model. In essence, AID is a heuristic approach to searching the raw data for structure and is purely a mechanization of the procedure a researcher might go through manually in hypothesizing a model.

Sonquist and Morgan (1964) hypothesized that the structure of the proper model would be indicated by the variables used in the splitting process and by the shape of the AID-Tree. They further hypothesized that symmetrical trees would be indicative of additive terms in the model which could be properly modeled by first-order predictors. More importantly, they felt that asymmetry, especially relative to the predictors involved in the splitting process, would be indicative of configural, or interactive, relationships among the variables. In this regard, they categorized tree structures as being either of the "trunk-twig" or of the "trunk-branch" type. The trunk-twig structure could be either of the top termination or of the bottom termination types. In the top termination type, small homogeneous subgroups (twigs) would be split off from the tree with each subgroup possessing some particular advantage. In the bottom terminating tree, the twig subgroups would contain data units which possessed alternative disadvantages. In the trunk-branch structure, the tree would be reasonably symmetric and would be reflective of cumulative advantages and disadvantages.

Sonquist (1970) presented the results of his continuing evaluation of the AID computer program as his doctoral dissertation. In this evaluation he was concerned with the analysis of synthetic data from three contrived models and the performance of the AID algorithm, vis-a-vis, the binary regression approach as embodied in the Multiple

Classification Analysis (MCA) program.[4] The data set consisted of seven variables divided into 49 categories, on each of 2036 data cases. The models used to generate the data were:

1) An additive model (trunk-branch)
2) A configural model in which advantages substitute for one-another but disadvantages are cumulative (upper terminating trunk-twig)
3) A configural model in which disadvantages substitute for one-another but advantages are cumulative (lower terminating trunk-twig)

Runs without noise, low noise ($\cong$ 10 percent), and high noise ($\cong$ 20 percent) were made and analyzed to see if they would result in the same models as were input.

The major conclusions of the investigation as presented by Sonquist (1970) were:

1) An additive model would produce a symmetric tree (Sonquist recognized that additivity was a sufficient condition for a symmetric tree but did not address the problem of necessity).
2) Small but universal effects may not be discovered in the splitting process.
3) Irrelevant variables which are correlated with powerful predictors may be used spuriously in the splitting process.
4) The presence of noise (at the levels considered) did not materially affect the resulting AID-Trees.

---

[4]MCA is described in Appendix A.

5) One evidence of interaction is a fairly large increase in the predictive power of one of the variables after the effects of some of the other predictors had been removed by the splitting process.

6) When a nonsymmetric structure appears, and the groups isolated first have approximately the same, low or high, means: a clear departure from additivity has occurred.

7) The AID-Tree was effective in depicting departures from additivity for the models considered.

8) The order in which the variables are used in the splitting process is not necessarily reflective of their overall importance in describing the process, rather the total variation explained by a variable may be more indicative of importance.

9) The β statistics calculated and output by the tested version of the AID program were virtually meaningless in depicting the relative importance of the individual variables.

Two important aspects of this investigation deserve special mention. The first is that the underlying models and the quality of the data set were known on an a priori basis. The second is that a data set of approximately 40 points per predictor category was available and used in the analysis.

It is entirely different problem to see a known model in the AID-Tree than it is to infer an unknown model from the AID-Tree. Further, the stability of the category means offered by a high data/predictor ratio helps significantly in discriminating between actual

effects that are due to the predictors and the spurious effects that are due to the idiosyncracies of the data sample.

Sonquist recognized these deficiencies in the generality of his results and suggested the application of the following heuristic procedures when analyzing AID-Trees.

1) The detailed profiles of each variable used in a split (and its competitors for the split) should be examined within the major subgroups throughout the tree.

2) All variables that are competitors in the splitting process should be examined closely for dependence. If two variables are measures of the same aspect of the situation, the variable that is least supportable purely on logical or theoretical grounds should be discarded.

3) The residual profiles of the final groups should be examined to see if effects from less powerful predictors may still be present and unaccounted for by the splitting process.

In regard to the size of the data set required for AID analysis, Sonquist (1970) stated:

> Data sets with a thousand or more cases are necessary, otherwise the power of the search process must be restricted drastically or those processes will carry one into a never-never land of idiosyncratic results. (p. 1)

In addition to generally confirming his original hypotheses relative to the utility of AID in detecting interaction effects, Sonquist's (1970) investigation did provide some insight into the problems to be

encountered in running AID. For these problems he offered the following general guidelines:

1) Do not categorize the variables with excessive detail; five to seven categories per predictor should be the maximum segmentation of the predictors.

2) Categorize the variables so as to minimize the problems created by skewness in the predictors: if a predictor is skewed, put those values in the tail of the distribution in separate categories.

3) Beware of skewness in the criterion; discard outlier cases or perform appropriate transformation of scale on them.

4) Perform appropriate initial runs with the split termination limits disabled in order to detect problems such as measurement or coding errors.

Finally, Sonquist (1967, 1970) proposed a procedure for using AID to hypothesize the terms of the model and then using MCA to determine the proper weights of the individual predictors. The additional analysis with MCA was espoused since the AID program does not produce statistics that accurately reflect the relative weights of the individual predictors.

Problems in the Application of AID for Hypothesizing Policy Models:

The application of this "structure searching" technique to the hypothesis of policy models would be straightforward except for two reservations--the amount of data required to obtain stable means for

the subgroups, and the lack of rules for inferring an "unknown" model from the data set.

Given the objective of "data conservatism", as presented in Chapter III, and the difficulties attendant in obtaining large quantities of decision data in operational environments, the freedom of the AID search would probably have to be somewhat restricted. Initial efforts by this investigator in analyzing the loan application data confirmed this fact. Splits resulting from groups of less than 10 cases were generally found to be meaningless and uninterpretable. In runs consisting of between 224 to 400 data units, with approximately 50 predictor categories, AID displayed a tendency to make spurious splits that did not hold up under cross-validation. When the splitting processes were restricted to avoid this overfitting, the explained variation in the data set was disappointingly low (< 50 percent).

The hypothesis of configural judgment models based solely on the splitting process was found to be less than successful, as measured by the lack of improvement in the predictive capability of the resulting configural models. The transformation from various AID-Tree structures into appropriate terms was not apparent. "Continuous" interaction terms of the form, $b_{ij}x_ix_j$, were found to be of little value in improving the fit of policy models. It was demonstrated that the number of categories and the definition of these categories could materially affect the appearance of the AID-Tree. Experimentation with various configural models revealed that perfectly symmetrical AID-Trees could be generated with data from highly configural models. Hence, additivity was found to be a sufficient condition for symmetry, but not a necessary condition.

Although the AID-Tree itself proved less interpretable than anticipated in inferring configural models for the decision process of the installment loan officers, it did retain the advantage of displaying the variables and their relationships in a manner that was readily comprehensible and believable to the judges involved. In all cases it identified those predictors that were most important to the judges. The predictors used in the early splits of the tree appeared to be pivotal elements in the judges' own descriptions of their thought processes.

In order to better define the proper configural relationships, it was concluded that the predictive efficiency and the correlation of the criterion with each of the predictors, within each subgroup of the AID-Tree, would be required in addition to the AID-Tree itself.

This same conclusion was independently reached by B. M. Finifter, of Michigan State University, who published the description of a proposed addition to the AID program in September, 1971. This addition consists of a routine to collect and display profiles of the Between Sums of Squares/Total Sum of Squares (BSS/TSS) from the AID splitting process. A normalized parameter reflective of the predictive ability of each predictor over the entire AID-Tree is calculated and designated as the Potential Explanatory Power (PEP) coefficient (Finifter, 1971, 1972).

The Development of the AID4UT/AIDTRE Computer Program:

The modification and extension proposed by Finifter was incorporated into an improved version of the AID4 computer program that was acquired from the Air Force Human Resources Laboratory.

The new program was implemented on the U. T. CDC-6600 computer system. It has the capacity to consider unlimited quantities of data with up to 80 predictors, divided into up to 700 predictor categories. The program can build, cross-validate, and double cross-validate AID-Trees, and it can produce profiles of the BSS/TSS for each predictor. A complete description of the modified program witn Users' instructions is provided in Appendix C.

Application of AID4UT/AIDTRE in Defining Subspaces:

Although the addition of the Finifter modification does provide more useful statistics for discerning unknown structure in the data, the problem of overfitting can only be avoided by limiting the number of splits that are allowed. In general, it was found that the limiting subgroups to contain no less than 5 percent of the cases would provide splits that were relatively stable. This early termination of the splitting process leaves considerable residual variability within the final subgroups. Review of these final subgroups revealed that more of the residual variability could be explained by various predictors, but the importance of any particular predictor was generally not as dominant as was the case in the earlier splits. This condition suggested that the residual variability in the subgroups might better be explained by a combination of the predictors instead of a binary split on a single predictor. In repeated applications on the 224 data units for the decisions of the installment loan officers, the criterion of retaining at least 5 percent of the total in each of the final subgroups

resulted in termination of the splitting process within 10 to 15 splits.[5] It was discovered that the most important predictors, in terms of their configural relationships with the other variables, tended to be involved in the early splits. Assessment of the subgroups resulting from these early splits indicated that they were often homogeneous. Within the individual subgroups, different combinations of the predictors appeared to explain the variation.

By relating this phenomena to the judges' own verbalizations of their policies, and reflecting upon the findings of Hoffman (1968), Bettman (1969), and Sonquist (1970), the concept of using AID-Trees solely for the purpose of defining a series of subspaces, and then modeling each subspace with a local model evolved. This technique was perceived as a means of taking advantage of the configural effects of the predominant variables without falling into the trap of overfitting the idiosyncracies of small subsets of data.

The Criteria for Defining Subspaces:

The decision of how many, and which subspaces to divide the predictor space into is based on heuristic rules developed during this approach. Experience with the loan officer data indicates the best criteria appear to be homogeneity of the subgroup and the predictive capability (both absolute and relative) of the predictors within the subgroups.

[5] Allowing the tree to split into smaller groups does not affect the first few splits since the tree can be manually truncated; this was the case with the loan application data.

The logic for defining appropriate subspaces is best explained by an example. The example that will be used is the definition of subspaces from the AID-Tree for one of the five installment loan officers whose policies were captured as a part of this research. An annotated AID-Tree for this policy model is shown in Figure 4-3; profiles for the BSS/TSS values for all 16 variables are shown in Tables 4-1 and 4-2. The data shown in Table 4-1 reflects the maximum between sum of squares for each predictor, in a given subgroup, normalized over the Total Sum of Squares for the entire data set. It will be designated as BSS/TSST. Table 4-2 reflects the profile in which the maximum between sum of squares for each predictor has been normalized over the Total Sum of Squares within the subgroup under consideration at a given step or trial. It will be designated as BSS/TSS(i). These two profiles reflect the "overall" and "local" predictive capability of each predictor. Comparison of these profiles are very helpful in revealing configural and nonlinear effects between the individual predictors even when one of the predictors involved is not actually used in the splitting process.

In the step-by-step consideration of the AID-Tree that follows, the superscripts refer to the corresponding entries on Tables 4-1 and 4-2. More detailed output for this analysis is shown in Exhibit 8 of Appendix B.

Step 1. Credit Rating has the largest value for BSS/TSST. Since there is only one group at this point, entries of .38[a, a1] appear in both Tables 4-1 and 4-2. (The next largest value is for Financial Reference which has .17[b]) Group 1 splits into groups 2 and 3. Group 2 is a homogeneous "twig" with a mean criterion value of .01. The input parameters used for this particular run of AID4UT[6] allow no

---

[6] The setting of the parameters for AID4UT is discussed in Appendix C.

ANNOTATED AID-TREE FOR EXAMPLE PROBLEM

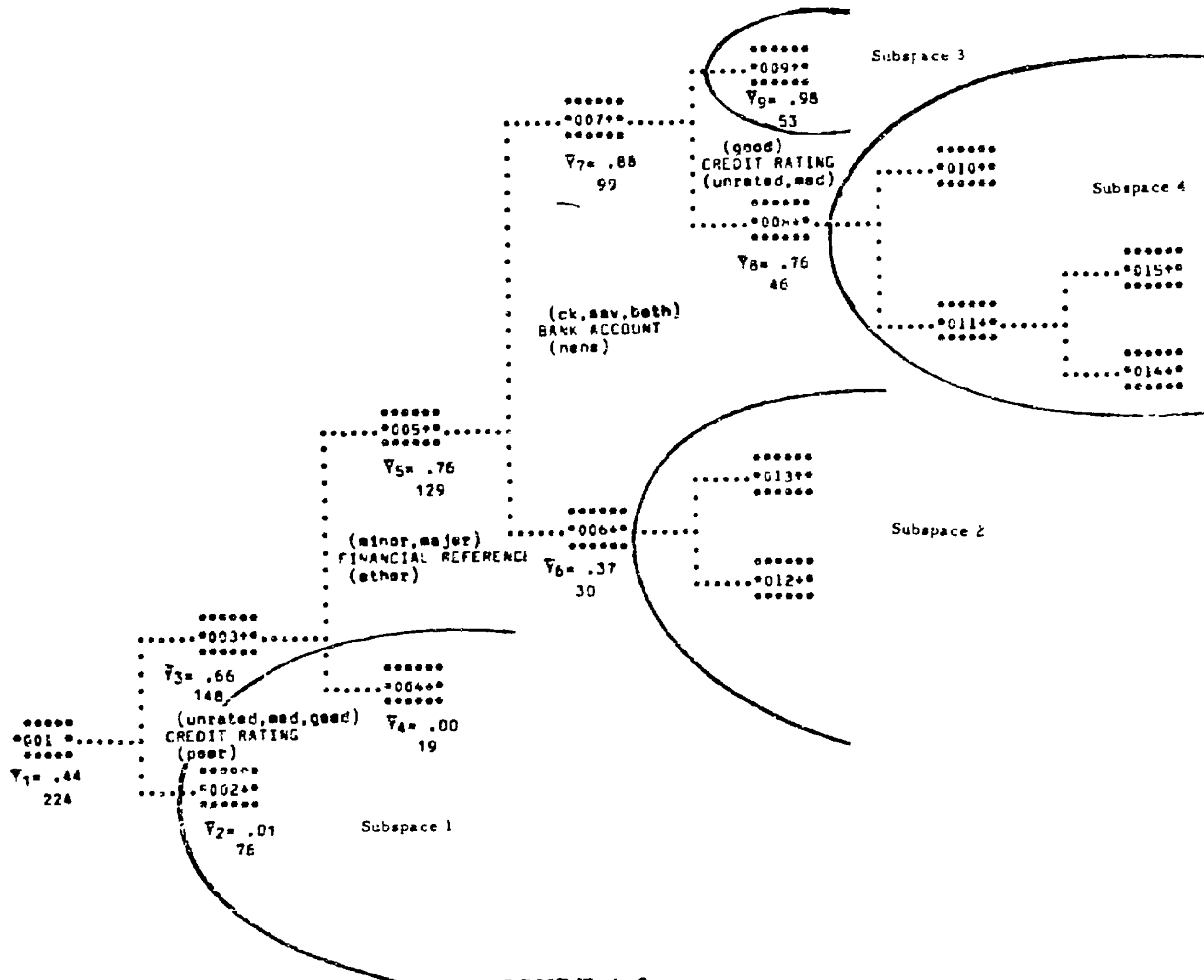


FIGURE 4-3

TABLE 4-1, BSS/TSST PROFILE

TRIAL/GRP		1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	1	.012652	.005742	.004474	.058844	.067264	.012138
2	3	.018847	.022938	.000285	.070939	.028827	.011871
3	5	.020705	.022297	.000030	.068832	.047718	.026762
4	7	.001245	.013164	.009270	.015649	.010720	.019124
5	8	.000791	.011929 m	.009838 n	.008500	.001518	.011832 k
6	6	.000259	.003017	0.000000	.026548 f	.001408	.060402
SUBSUM		.05450	.07909	.02390	.24931	.15746	.08213
TRIALS		6	6	6	6	6	5
MEANBS		.00908	.01318	.00398	.04155	.02624	.01369
S D		.00867	.00751	.00424	.02548	.02456	.00802
CF VAR		.95475	.56985	1.06340	.61320	.93595	.58607
VAR		.000[illegible]	.00006	.00002	.00065	.00060	.00006
RECSUM		.05450	.07909	.02390	.24931	.15746	.08213

TRIAL/GRP		7RESCAT	8TYPEMP	9INCOME	10BANKAC	11FINREF	12LOANAM
1	1	.055871	.043629	.032889	.124566	.171450 b	.022689
2	3	.066715	.022979	.027645	.118880	.173004 c	.009053
3	5	.036919	.028279	.025921	.109299 e	.052360	.009254
4	7	.014633	.008202	.013087	.000911	.005730	.003632
5	8	.010245 j	.004477	.005348	.001797	.005985	.013854 i
6	6	0.000000	.003510	.011533	0.000000	.002428	.004927
SUBSUM		.20438	.11108	.11616	.35545	.41096	.06362
TRIALS		6	6	6	5	6	6
MEANBS		.03[illegible]	.01851	.01936	.07109	.06849	.01057
S D		.02636	.01457	.00999	.05715	.07527	.00635
CF VAR		.77385	.78721	.51610	.80389	1.09896	.60041
VAR		.00069	.00021	.00010	.00327	.00567	.00004
RECSUM		.20438	.11108	.11616	.42654	.41096	.06342

TRIAL/GRP		13LOANTM	14NECESS	15EQUITY	16CRERAT
1	1	.004920	.037048	.001647	.382847 a
2	3	.005929	.017527	.000234	.112233 d
3	5	.000763	.004135	.000022	.091735
4	7	.000645	.001469	.000091	.021626 h
5	8	.001543	.001916	.000052	.000007
6	6	.001821	.000014	.005430	.040882 g
SUBSUM		.01562	.06211	.00753	.64933
TRIALS		6	6	6	6
MEANBS		.00260	.01035	.00125	.10822
S D		.00206	.01329	.00196	.12874
CF VAR		.78993	1.28386	1.56027	1.18964
VAR		.00000	.00018	.00000	.01658
RECSUM		.01562	.06211	.00753	.64933

SUMMARY TABLE

		RECONSTRUCTED SUMBSS/TSST	RANK
1	AGE	.05450	13.0
2	MARITL	.07909	10.0
3	LOCAL	.02390	14.0
4	PHONE	.24931	4.0
5	JOBTIM	.15746	6.0
6	RESTIM	.08213	9.0
7	RESCAT	.20438	5.0
8	TYPEMP	.11108	8.0
9	INCOME	.11616	7.0
10	BANKAC	.42654	2.0
11	FINREF	.41096	3.0
12	LOANAM	.06342	11.0
13	LOANTM	.01562	15.0
14	NECESS	.06211	12.0
15	EQUITY	.00753	16.0
16	CRERAT	.64933	1.0

TABLE 4-2, BSS/TSS(i) PROFILE

TRIAL/GRP		1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	1	.012652	.005742	.004474	.058844	.067264	.012138
2	3	.031450	.038275	.000476	.118372	.048102	.019809
3	5	.048572	.052305	.000071	.161468	.111939	.062779
4	7	.006522	.068966	.048553	.081984 h3	.056162	.100187 h2
5	8	.005218	.078742 m1	.064935 m1	.056104	.010017	.078099 k1
6	6	.002051	.023923	0.000000	.210526 f1	.011164	.003190
SUBSUM		.10646	.26795	.11852	.68730	.30465	.27620
TRIALS		6	6	6	6	6	6
MEANBS		.01774	.04466	.01975	.11455	.05078	.04603
S D		.01680	.02515	.02663	.05628	.03483	.03632
CF VAR		.94675	.56324	1.34800	.49135	.68599	.78892
VAR		.00028	.00063	.00071	.00317	.00121	.00132
RECSUM		.10646	.26795	.11852	.68730	.30465	.27620

TRIAL/GRP		7RESCAT	8TYPEMP	9INCOME	10BANKAC	11FINREF	12LOANAM
1	1	.055871	.043629	.032889	.124566	.171450 b1	.022589
2	3	.111324	.038343	.046130	.198367	.288682 c1	.015106
3	5	.133523	.066338	.060806	.256398 e1	.122828	.021732
4	7	.075660	.042971	.068550	.004772	.030017	.019025
5	8	.067625 j1	.029553	.033596	.011864	.039509	.091450 i1
6	6	0.000000	.027838	.091453	0.000000	.019253	.039075
SUBSUM		.44500	.24867	.33343	.59597	.67174	.20909
TRIALS		6	6	6	5	6	6
MEANBS		.07417	.04145	.05557	.11919	.11196	.03485
S D		.04240	.01267	.02070	.09973	.09613	.02640
CF VAR		.57173	.30567	.37247	.83674	.85864	.75761
VAR		.00180	.00016	.00043	.00995	.00924	.00070
RECSUM		.44500	.24867	.33343	.71516	.67174	.20909

TRIAL/GRP		13LOANTM	14NECESS	15EQUITY	16CRERAT
1	1	.004920	.037048	.001647	.382847 a1
2	3	.009893	.024246	.000391	.197277 d1
3	5	.001791	.004701	.000053	.215195
4	7	.003382	.007697	.000475	.113296 h1
5	8	.010100	.012650	.000346	.000047
6	6	.014440	.000109	.043062	.324191 g1
SUBSUM		.04461	.09645	.04602	1.22285
TRIALS		6	6	6	6
MEANBS		.00744	.01608	.00767	.20381
S D		.00442	.01285	.01584	.12693
CF VAR		.59464	.79941	2.06449	.62277
VAR		.00002	.00017	.00025	.01611
RECSUM		.04461	.09645	.04602	1.22285

SUMMARY TABLE

		RECONSTRUCTED SUMBSS/TSS(I)	RANK
1	AGE	.10646	13.0
2	MARITL	.26795	9.0
3	LOCAL	.11852	12.0
4	PHONE	.68730	3.0
5	JOBTIM	.30465	7.0
6	RESTIM	.27620	8.0
7	RESCAT	.44500	5.0
8	TYPEMP	.24867	10.0
9	INCOME	.33343	6.0
10	BANKAC	.71516	2.0
11	FINREF	.67174	4.0
12	LOANAM	.20908	11.0
13	LOANTM	.04461	16.0
14	NECESS	.09645	14.0
15	EQUITY	.04602	15.0
16	CRERAT	1.22285	1.0

further splitting of this group because none of the predictors can explain more than .01 of the total variability.

Step 2. In group 3, Financial Reference has the largest BSS/TSST of .17[c] in Table 4-1 and a BSS/TSS(i) of .28[cl] in Table 4-2. Note that the BSS/TSST remained the same for this variable but the BSS/TSS(i) increased markedly. This would indicate that Financial Reference is immaterial at a Credit Rating level of "poor", but that it is the single most important variable over the Credit Rating levels of "unrated", "medium", or "good". Further, note that the predictive capability of Credit Rating [d1] decreased by a factor of two on the "local" (BSS/TSS(i)) basis after the split. This would indicate that there is a much smaller difference between the upper three levels of Credit Rating than there is between those three levels and the lowest level, "poor". By splitting group 3 on Financial Reference, a second homogeneous "twig", group 4, is formed consisting of 19 data units which have a mean criterion value of .00. Because of the similarity of their mean criterion values, group 2 and group 4 are combined together into subspace 1. A one term model (a binary indicator of group membership) is sufficient for this subspace.

Step 3. In group 5, Bank Account has a BSS/TSST of .11[e] and a BSS/TSS(i) of .25[e1]. It is the basis for splitting group 5 into subgroups 6 and 7. Neither of these resultant subgroups is as homogeneous as were the previous "twigs". Each will require further modeling. Considering the BSS/TSS(i) profiles for group 6, note that both Phone [f, f1] and Credit Rating[g, g1] show sizable values. However, the group 6 contains only 30 data units and a binary split on any one predictor at this level will result in groups that contain the minimum allowable number of data units and are still nonhomogeneous. At this

point, it is better to model group 6 with a first-order model containing both Credit Rating and Phone. Thus, group 6 becomes subspace 2.

Step 4. In group 7, Credit Rating has a BSS/TSS(i) value of .10$^{(h1)}$ which is only slightly larger than the values for Restime$^{(h2)}$ and Phone$^{(h3)}$. However, if allowed to split on Credit Rating, a homogeneous "twig" will result. Group 9 has 53 data units and a mean criterion value of .98. It comprises subspace 3 and is modeled by a one term model.

Step 5. In group 8, Loan Amount$^{(i, i1)}$, Rescat$^{(j, j1)}$, Restime$^{(k, k1)}$, Marital$^{(m, m1)}$, and Local$^{(n, n1)}$, all have BSS/TSS values reflecting predictive capability, but a binary split on any single one of these variables will not result in a homogeneous group. This group becomes subspace 4.

This walk-through of the heuristic logic employed in defining subspaces in an AID-Tree brings out two points that were observed during the analysis on the various AID-Trees in this research. First, reasonably sized groups which are homogeneous are prime candidates as subspaces, especially if they are split off in the very early stages of the AID-Tree. Second, in those parts of the AID-Tree where the groups are not homogeneous, predictors that cannot explain in excess of 10 percent of the variability within the subgroup will probably not result in isolation of new homogeneous subgroups of substantial size: in this case it appears better to fit local "continuous" linear models.

Building Local Models Within the Subspaces:

Once the subspaces are identified, the data in the BSS/TSS(i) profile, Table 4-2, can be of further use in hypothesizing the proper

terms for the models. In formulating the models for the individual subspaces, there are several possible approaches. If there is sufficient data available (in cases where at least one data unit per predictor category is available, and preferably on the order of 5-10 data units per predictor category), a binary regression model could be built by using either the MCA program or any standard regression package. In subspaces where the number of data units is not amenable to using binary or categorical models, the formulation of a "continuous" first-order regression model is usually a reasonable alternative. Under the latter formulation, the investigator effectively imposes the assumption that a unit change in the predictor value will produce a fixed (constant) change in the criterion variable. He makes this assumption in exchange for using fewer degrees of freedom and predictors in the model.

In the definition of subspaces and local models for the five loan officers, various alternative formulations were attempted at the subspace level. For example, the approach of combining two neighboring subspaces into a larger subspace and formulating a configural model over this larger subspace was compared with the approach of formulating separate first-order models over the original subspaces. The general conclusion drawn from these efforts was that the approach of using first-order models over the individual subspaces was superior.

The Determination and Evaluation of Model Coefficients:

The task of obtaining regression coefficients for all of the models investigated was performed using an iterative determination/cross-validation procedure. A similar procedure was originally suggested by a member of the research supervising committee,

Dr. E. Jennings, as a method of evaluating the stability of alternative models. The technique consists of the random selection of 80 percent of the data units from the input file, generation of the coefficients using the REGREJ routine of the EDSTAT-J system (Jennings, 1971), and cross-validation of these coefficients with the remaining 20 percent of the data units. This cycle was accomplished 10 times and resulted in a profile of coefficients for each predictor. Whereas Jennings uses this procedure solely to assess the stability of various models, in this research, the average values for each of the coefficients were actually used in the model. The objectives of this revised procedure were the simultaneous assessment of the stability of the model and the reduction of the idiosyncratic effect that any outlier data units would have on the model coefficients. The average coefficients were compared with the corresponding coefficients obtained in a single fit over all 224 data units for each of the five judges. They were found to be very similar. One interesting outcome of this comparison is reflected in Table 4-3. Note that the average values for predictive efficiency (R^2) for the 10 sets of coefficients, resulting from the iterative scheme, are slightly better than the corresponding values of R^2 that were obtained in the single fit procedure. This would suggest that, indeed, some of the idiosyncratic effects of outlier points may have been overcome by the iterative procedure.

TABLE 4-3

COMPARISON OF PREDICTIVE EFFICIENCIES

FOR ALTERNATIVE COEFFICIENT DETERMINATION PROCEDURES

Procedure	Judge 1	2	3	4	5
Iterative Procedure	.670	.707	.680	.633	.632
Single Fit Procedure	.659	.695	.680	.614	.618

Comparative Results of Various Modeling Approaches:

Installment Loan Officer Models

The quality of the various models investigated in this research were initially measured in terms of the standard statistical parameter, R^2, and the cross-validity from one subset of the 224 sample points to another subset. An 80/20 split between the determination sample and the cross-validation sample was used. The average cross-validity for 10 trials per model of the first-order, second-order, and local models are shown in Table 4-4; the data for the AID splitting process consists of only one trial per AID-Tree.[7] These data reflect

[7] The process of using an 80/20 split in the AID cross-validation process was found to influence the results in the cases of limited data sets since the groups in the subgroups propagated in the 20 percent sample become very small and susceptible to group means that were greatly affected by random sampling error. The cross-validity coefficient in AID is not generally monotonic and the values presented in Table 4-4 reflect the maximum cross-validity attained in the forced splitting.

TABLE 4-4

COMPARISON OF CROSS-VALIDITY ON SPLIT SAMPLE

Model	Judge 1	2	3	4	5
First Order	.579	.624	.634	.512	.538
Second Order	.649	.645	.653	----	----
AID-Split	.440	.621	.675	.670	.547
Local Model	.727	.659	.685	.633	.558

that the local models, derived from the AID-Tree, were consistently better than the first-order or the second-order models.

In actually evaluating the goodness of the various modeling approaches in predicting dichotomous decisions of the installment loan officers hit rate is a more relevant measure of goodness than R^2. Since each of the models resulted in a predicted score in the normal range from .00 to 1.0,[8] each possible value in this range could be used as the cut score, or discrimination point. Above this point the value of "1"(Yes) would be assigned to the data unit, and below this point, a value of "0" (No) would be given. Generally, at each possible cut point there are some data units that will be misclassified (either actual 0's given the value of 1, or actual 1's given the value of 0). The hit rate reflects the percentage of proper classifications, but does not reflect the type of misclassification that is being made. Data on cross-validation hit rates was commensurate with that shown in Table 4-4.

[8] Actually scores slightly smaller than .00 and slightly larger than 1.00 can be obtained but in those cases the scores are rounded to .00 or 1.00 respectively.

The indifference of the magnitude of hit rate to type of error is extremely important since the decision maker may be more concerned with one type of error than he is with the other type. Hence, the performance of the model over a range of possible cut points is important. Figure 4-4 schematically depicts the profiles of hit rates for first-order

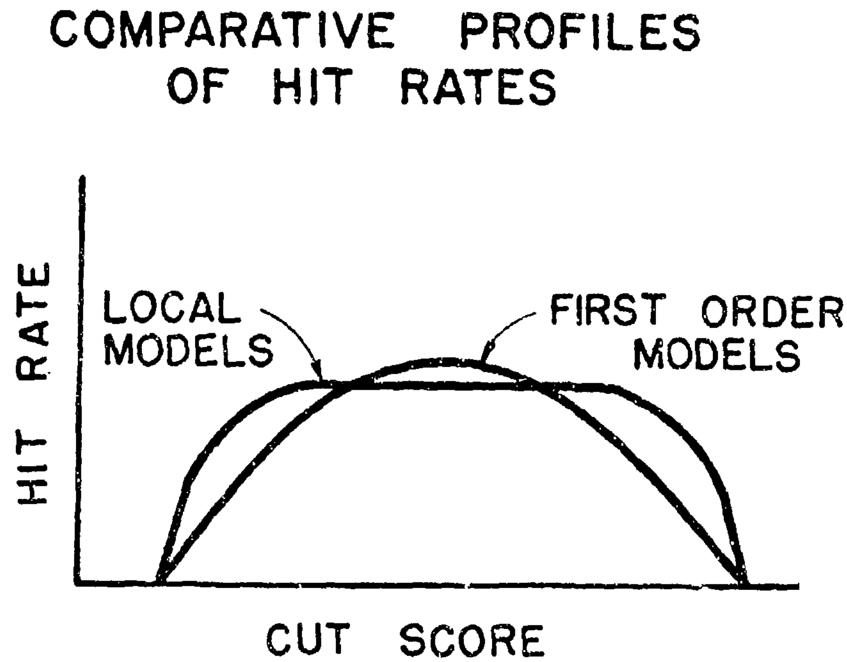


FIGURE 4-4

and local models for the installment loan officer data. It shows that even if the maximum hit rate of the first-order model is marginally better, the profile is generally inferior and the marginal superiority prevails over only a very small range of cut scores. Thus, the single parameter embodied in either the hit rate or the cross-validation coefficient is probably not the best criterion for determining overall superiority, but rather, its magnitude and profile should be considered

jointly. Figures 6-14 through 6-18 in Chapter VI reflect that both the maximum hit rates and the profiles were superior for the local models in the 80/20 cross-validation process.

A completely separate cross-validation effort using 150 data units collected during a different time period was also performed. The maximum hit rates for the first-order models were slightly superior to those for the local models in this test. However, this reversal can be traced to a significant shift in the characteristics of the data set and a shift in the policy of the judges. The full details of this test are discussed in Chapter VI.

The Valenzi Data, an Independent Example

Whereas the approach of defining subspaces and local models was conceived as a result of the initial analysis of the policies of the loan officers and was developed primarily using this data, it might be criticized as being useful only in situations with binary criterion values. An insight into the generality of the local modeling approach was sought by applying it to an independent[9] set of data for which a continuous criterion existed. The data used for this analysis was published by Valenzi (1970). The data consisted of the judgments made by employment counselors relative to the chances of applicants being hired for secretarial jobs. Five tricotomous predictors (typing speed, shorthand speed, experience, education, and social skills) were used to describe each applicant. The criterion was a scalar value from 0 to

[9]Independent in this context denotes that the investigator had not previously analyzed the data and was not able to intuit a good model as a result of familiarity with the process or the data.

10. A full 3^5 factorial experimental design with two replications per cell was analyzed with both ANOVA and multiple regression models. Valenzi attempted to demonstrate and evaluate configural cue use by four judges; the analysis and comparisons in the following discussion considers only one judge, Judge 4.

Valenzi published the following results for Judge 4:

1. ANOVA results with main effects and all two-way interactions resulted in 73 percent of the variability being explained by the main effects and 14.5 percent being explained by the second-order configural effects. A total of 87.5 percent of the variability was explained with some 41.5 percent being explained by the educational main effect.
2. The second-order multiple regression model resulted in 60 percent of the variation being explained by the first-order terms and 9.7 percent being explained by the second-order terms.
3. The largest interaction identified by either of these processes was Education by Social Skills which explained 8.5 percent of the variability in Valenzi's ANOVA model and 5.4 percent in his second-order regression model.
4. The importance ranking for the predictors given by Judge 4 did not match the rankings derived from the statistical analysis models.

Valenzi concluded that Judge 4 did exhibit configural cue use and that ANOVA was a better approach to detecting it than were

standard multiple regression techniques or the "a priori partialling out" of the configural variance suggested by the judge's verbalization of his policy.

The results obtained by applying the AID4UT/AIDTRE computer programs and the local modeling approach to the same data were:

1. An AID-Tree with 14 final groups explained 89 percent of the variability. Using an 80/20 split, a cross-validity of 85 percent was achieved.

2. The first split in the AID-Tree was made between the lowest and the upper two levels of Education. This split explained 40.1 percent of the variability and resulted in a homogeneous "twig" which included 162 data units (1/3 of the total data set). This group had an average criterion value of .01 on a scale from 0 to 10.

3. The most important variable for the other 324 cases was social skills. Its interaction with all other variables was readily detectable in the detailed output from AID4UT.

4. Definition of two subspaces on the basis of the first split and the use of the following local models. (Subspace 1 included all cases with the lowest level of Education and subspace 2 included all others.)

   a. $Y = b_0 + \bar{Y} + e$: for subspace 1

   b. $Y = b_0 + b_1$ (social skill)(typing)

      $+ b_2$ (social skill)(shorthand)

      $+ b_3$ (social skill)(education)

$+ b_4$ (social skill)(experience) $+ e$:

for subspace 2

resulted in a 6 predictor model which explained 81.8 percent of the variance, a gain of 12.1 percent relative to the 15 predictor second-order regression model used by Valenzi.

5. Changing the model over subspace 2 to a binary regression model which considered all two-way interactions, 90 percent of the variation was explained.

6. The rank order of the variables as determined by the profile of BSS/TSS(i) from the AID4UT program was more consistent with the judge's verbalizations than the corresponding ranking achieved by Valenzi with ANOVA and regression analysis. Table 4-5 reflects the comparison. A complete reversal of the ranks of the first four variables arose from ANOVA. Only social skills was out of sequence with AID4UT.

TABLE 4-5

COMPARISON OF RANKED IMPORTANCE OF PREDICTORS

Rank	Judge's Verbalization	ANOVA	AID4UT
1	shorthand	education	social skills
2	typing	social skills	shorthand
3	social skills	typing	typing
4	education	shorthand	education
5	experience	experience	experience

The implications of the comparison of the results achieved on the Valenzi data are clear. The use of the AID algorithm to isolate homogeneous subgroups over which better fitting local models can be applied is potentially superior in situations other than those with dichotomous criteria. The superiority of the technique is not solely a function of having "lived" with the data and the decision process for a prolonged period of time. In this independent example, the model achieved was simpler and easier to reconcile with the verbalized policy of the judge.

Summary of Guidelines for Applying AID4UT/AIDTRE in Policy Capturing:

The experience gained thus far in applying AID4UT/AIDTRE to Policy Capturing is extensive in the sense that numerous models have been attempted and produced, but limited in the sense that these models were all for a particular decision process, except for the brief demonstration on the Valenzi data.

Table 4-6 presents a set of guidelines that should make future applications of the AID algorithm and the local modeling technique more efficient and effective. These guidelines reflect the experience of Sonquist (1970) as well as the experience gained in the current research effort. They are directed at using AID4UT/AIDTRE for the definition of subspaces and local models when there is limited data available.

These empirical guidelines for applying AID are somewhere between the stages of artistic and heuristic. This investigator doubts whether such guidelines will become more than heuristic in the foreseeable future. Considering that AID is a model seeking technique and its functions is to uncover relationships in data sets with unknown

structure, it is improbable that precise, deterministic criteria for defining subspaces or for specifying exactly which predictors are important, can be defined. If enough information were known about the data set to do this, the very need for the model seeking technique would be doubtful.

TABLE 4-6

GUIDELINES FOR EMPLOYING AID4UT/AIDTRE
IN BUILDING POLICY MODELS

PREPARATION OF DATA:

1. Segment predictors into 3-8 categories. A lower limit of at least 5 data units per category should be sought.

2. If the criterion variable is badly skewed either divide the data set into subgroups or make a transformation on the criterion such as $Y' = \log Y$.

3. Obtain a contingency table for the predictors. An association measure such as the Goodman and Kruskal λ as produced by Anderberg's (1971) GCORR program is helpful. Note those variables with very high associations since they may be measures of the same phenomena and substitute for each otherin the splitting process. In such cases one of the variables should be eliminated from consideration as a predictor.

INITIAL RUNS OF AID4UT/AIDTRE

1. Run AID4UT with all stopping parameters disabled. (See Appendix C). This produces a "configural" model which provides an upper limit on the capability of the set of predictors under consideration to predict the variability in the data. For large data sets with many predictor categories, this run may be very time consuming and setting the minimum subgroup size to .01 of the total set may be advisable.

2. Refine/collapse the predictor categories if possible. If two or

more categories of a predictor never end up in different subgroups as a result of a split they can be combined into one category.

3. Analyze the residuals for obvious coding errors. Such errors are usually subtle, but can affect later results considerably.

ANALYSIS RUNS OF AID4UT/AIDTRE

1. After corrections and modifications from the initial runs are completed, rerun the analysis with increased stopping parameters. In data limited cases where less than 10-20 data units per predictor category exist, set split reduceability $\geq .01$ and the minimum group size to $> .05$ of the original sample size.

2. Analyze tree to identify large, homogeneous "twig" groups indicated by:

   a. Mean criterion values of $> .9$ or $< .1$ for dichotomous criteria
   b. Standard deviations of the criterion values of $< .2$ of the standard deviation of the original group for continuous criteria.

3. Study the mean criterion values for each predictor category in each subgroup for the first 4 or 5 splits. Pertinent points are:

   a. The difference in the magnitude of the mean criterion values in two subgroups resulting from a split reflect the "main effect" of the split variable over the parent group. Of particular importance are the trends of the mean criterion values throughout the AID-Tree. Plots of the mean criterion values versus the predictor categories for each of the subgroups in the AID-Tree will portray these trends If the trends are similar then the predictor is likely to be independent of the variables used in the previous splits; if the trends for a predictor vary greatly, there may be interaction present between the predictor and the variables used in the previous splits.
   b. If the standard deviation of the criterion values for a given predictor category is near zero and the category is small ($< .05$ of total sample) this represents a homogeneous

group which may not be automatically identified by the AID-Tree because of its size and the intervention of more powerful predictors in the early stages of the splitting process. If there is a plausible theoretical reason for this group it could be removed from the data set in future analysis.

4. Analyze the BSS/TSS profiles to ascertain predictive capability of the individual predictors over the AID-Tree. Pertinent points are:

   a. If BSS/TSS(i) is greatly (a factor of 2) different in the groups resulting from a split, two situations may exist; either there is an interaction between this variable and the split variable or they are measures of the same phenomena. A check of the contingency table would indicate if they are highly correlated and if substitution is likely. If the BSS/TSS(i) for the split variable changes, either groups with a single category of that predictor have been created or there is a higher order effect for the predictor.
   b. If the BSS/TSS(i) value for a predictor remains the same after it has been involved in a split, but it loses the next split, a "continuous" interaction (as opposed to the discrete interactions represented by discrete hyperplanes) may be present. Separate analyses for each of the categories of the predictor involved would result in similar AID-Trees for each category, however, the mean criterion response values for corresponding subgroups in the various AID-Trees would be different.
   c. If the BSS/TSS(i) for a predictor remains small ($<.05$), but relative constant over the entire AID-Tree, the predictor may have a small universal effect and could be included in each local model.
   d. Comparison of the BSS/TSS(i) and BSS/TSST profiles will give further indications if a variable has a universal or only a local effect. Those variables which appear very low in importance on both scales can probably be discarded from the predictor set.
   e. The identification of important variables for the various subgroups can be accomplished from the profile of BSS/TSS(i). Listing the most important variables on the skeleton version of the AID-Tree will provide a visual impression of groups in which the same predictors a e effective.

DEFINITION OF SUBSPACE AND LOCAL MODELS

1. Subspaces can be identified on three bases.

    a. Homogeneity relative to the criterion value.
    b. Consistency relative to the predictors that are important within the subgroup.
    c. Consistency relative to the effects that the individual predictors have in the local subgroups.

2. If there are no homogeneous groups identifiable on the basis of the above characteristics, the total predictor space should be represented with one "continuous" model. In this case the hypothesis of "continuous" interaction terms of the type, $b_{ij}x_ix_j$, can be accomplished by considering the BSS/TSST profile.

3. If there are homogeneous subspaces within the AID-Tree, the information from the splitting process provides the basis for the generation of these subspaces directly. The residuals list may be used to identify and physically separate the data cases in the various subspaces for separate analyses. Alternatively, the logical capabilities of the computer can be used to define binary subspace multipliers. CAUTION: It is extremely easy to define overlapping subspaces if the split process of the AID-Tree is not followed precisely.

4. Once the subspaces are identified, the predictors within each local model are hypothesized on the basis of the BSS/TSS(i) profile. If sufficient data units exist, binary models can be formulated. Otherwise, "continuous" first-order regression models appear to be the best approach. Those variables not inherently monotonic should be checked for this property over the subspace under consideration. This is an implicit assumption of "continuous" formulation and the BSS/TSS values do not necessarily reflect this restriction.

GENERATION OF REGRESSION COEFFICIENTS

Any available regression package can be used to generate the appropriate regression coefficients. The data transformation capabilities available in the DATRAN feature of the EDSTAT-J package are

particularly compatible with the logical definition of subspaces on the computer. The use of the iterative coefficient determination/cross-validation approach is recommended since it allows assessment of the coefficient stability in the process of model building.

CHAPTER V: BACKGROUND, ENVIRONMENT, AND CHRONOLOGY OF CASE STUDY

Introduction:

This chapter provides the scenario and a brief history of the activities of this research that were specifically related to capturing and evaluating the policies of a group of installment loan officers. The results and implications are then presented in Chapter VI.

A Brief Resume of Numerical Credit Scoring:

The application of numerical rating systems to credit evaluation was first proposed by Durand (1941) in a study of several hundred accounts. He used discriminant analysis to categorize those accounts on the books as "good" and "bad" on an "ex-post" quality[1] basis. His study, and others that followed, generally indicated the potential utility of numerical credit scoring systems in screening out bad accounts. However, there was no meaningful implementation of any such systems until about 1960. Myers and Forgy (1963) attribute this lack of implementation to:

1) a natural reluctance on the part of the experienced credit executive to abandon the time honored "judgmental" approach in favor of the newer and relatively untested quantitative methods;
2) the inability of statisticians to develop "foolproof"

[1] "ex-post" quality is based on collection experience and "ex-ante" quality is based on prospective risk. The decision process considered in this research is based on the "ex-ante" quality of the loan application.

rating systems which consistently identify poor credit risks accurately enough to result in substantial net savings by the credit operation;

3) the difficulties involved in utilizing any such effective rating system in the operating situation; and

4) unwillingness on the part of the statistical consultants to invade the domain of the credit manager and do the selling job necessary to transform such an idea into a successful and useful operating tool. (p. 120)

In their own study, Myers and Forgy (1963) set about analyzing various statistical modeling techniques for scoring loans on mobile homes. Again they used the common procedure of scoring "ex-post" quality and did not consider the judgment process of the credit executive nor did they demonstrate the operational feasibility of the system. They did make one advancement by validating their models by means of cross-validation.

The expanding availability and use of computers within the credit industry during the mid-1960's propagated renewed interest in numerical scoring systems and considerable debate as to their utility has transpired in financial journals such as Credit World and Banker's Monthly. Continued research in the area has proceeded under the sponsorship of the National Bureau of Economic Research, the Credit Research Foundation, the First Pennsylvania Bank and Trust, and Morris Plan Company of California.

By 1968 some 35% of the top 200 banks in the United States were using numerical credit scoring of some form, and another 33% were considering such systems. Of those banks having systems, some 80%

had been borrowed by one institution from another. The borrowing of scoring systems was attributed to the extensive analytical effort involved in setting up a "tailored" system, and was responsible for significant implementation problems. (Wilt and Tierney, 1968) As a result of these problems brought about by using borrowed systems, the credibility and utility of numerical credit scoring has suffered even more in the eyes of the critics.

In a review of the first decade of use of numerical scroing systems, Zaegel (1971) noted that the benefits from use of such systems could not be accurately measured, but that losses were down approximately 20% in situations where the systems were being used. He did note that many improvements could be made especially with regard to the time it took to establish and validate a system, which requires about five years for the loan turnover cycle when the system is based on "ex-post" quality.

Although brief, this resume sets the stage for the research into applying Policy Capturing techniques in defining credit scoring models that are more "personal", "flexible", "understandable", and hence, more likely to be implemented in operational environments.

The Decision Process Considered in this Study:

The primary decision process modeled in this research is that of the installment credit officer in making judgments of whether to grant or to deny a loan to an applicant based on information contained in a written application. The decisions being modeled are limited to those involving "dealer paper", i.e., loan applications forwarded to the lending institution as a result of a dealer contract. Under a dealer

contract, the lending institution provides the financing for a merchant's sale on an individual account basis, i.e., they agree to buy the installment contracts from a merchant on an individual basis if the applicant and the account meet the lending institution's standards of acceptability. Unlike more general credit decisions in which the interest rates and other particulars of the loan are negotiable parts of the decision, the dealer contracts under consideration carry a fixed interest rate and provide that the merchandise will serve as collateral. Further, the amount and duration of the loan are not actually considered as items of negotiation, but as parameters of the loan. Thus, the decision is essentially uni-dimensional in that only a binary decision of "accept" or "reject" is made.[2]

A flow diagram of the "dealer paper" process is shown in Figure 5-1. Note that no personal contact exists between the loan applicant and the credit officer and the decision is made solely upon the basis of the input data contained on the written application and the applicant's credit rating that is obtained from the local credit bureau.

The original application form (Exhibit 1, Appendix B) is filled out by the clerical staff at the dealer location and is then telephoned to the lending institution. There the clerical staff of the installment loan department transcribes the information and calls the local credit bureau for a credit report. Upon receipt of this report, the information is attached to the application and forwarded to an available loan officer for his judgment. If he is a senior loan officer

[2] In special cases an "accept with full recourse" is made in which the lender grants a loan he would ordinarily reject if the merchant guarantees payment in case of default. In this study, these are considered rejects.

DEALER PAPER FLOW DIAGRAM

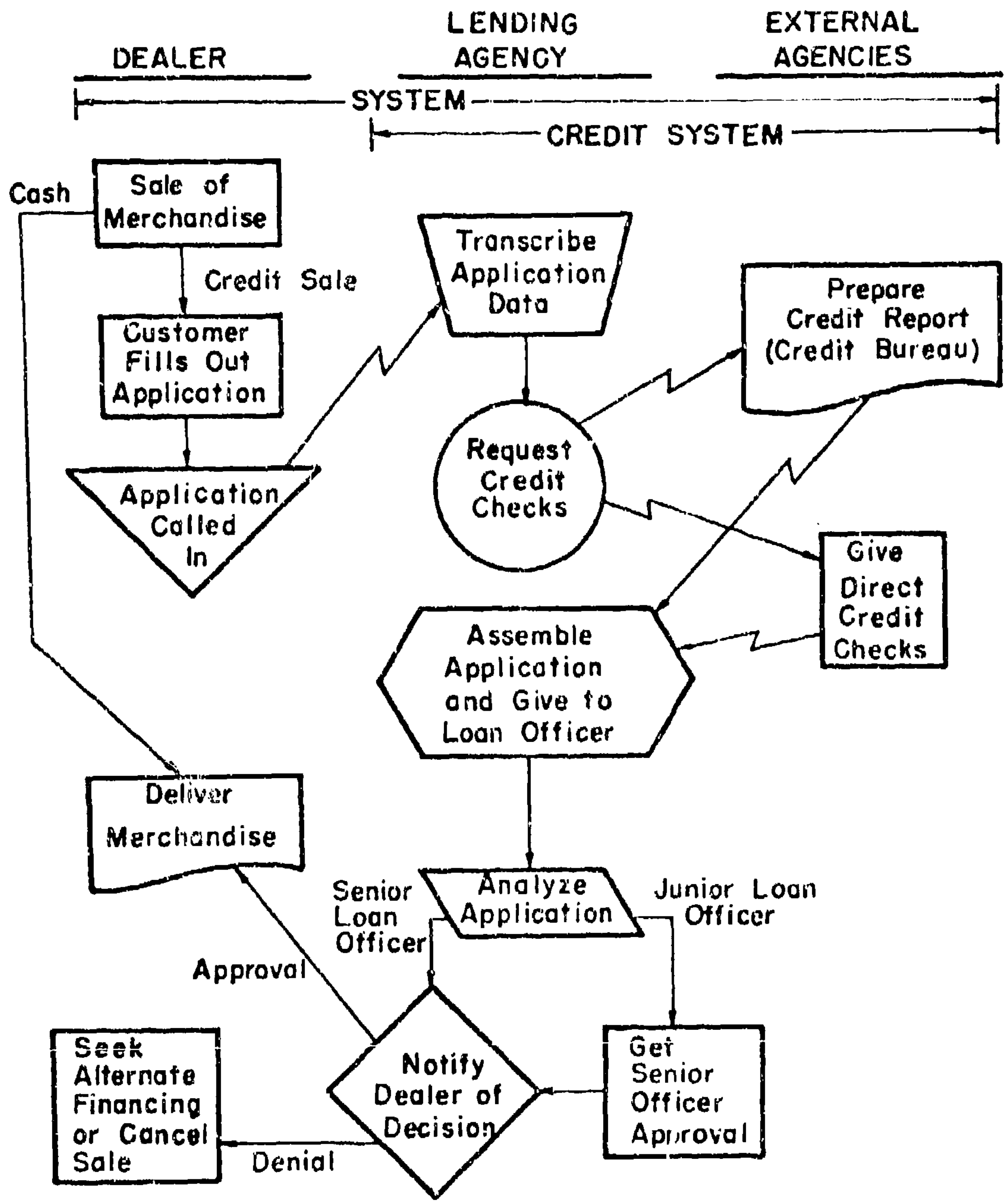


FIGURE 5-1

his judgment is generally conclusive, but if he is junior, he normally makes a recommendation and obtains concurrence from one of the senior loan officers. When the decision is made, the dealer is notified. If the loan is accepted, he executes the contract and releases the merchandise to the customer; at this point the responsibility for maintenance of the account and collection of the loan devolves to the lender. If the decision is denial, the dealer must either seek financing elsewhere, carry the note himself, or vitiate the sale.

The entire process, from sale to a decision on the loan, involves approximately four clerks, two transcriptions of the data, and the judgment of one or more loan officers. It requires from one hour to several days. For purposes of later discussion, the "system" will include those elements from sale until notification of decision and the "credit system" will include those elements from receipt of the application by the lending institution until the ratification of the decision.

The Lending Institution:

The installment credit department participating in this study is part of Lamar Savings and Loan Association located in Austin, Texas. The department is undergoing a transition period of rapid growth. In the last three years their outstanding installment accounts have increased over 500% from less than $2 million to over $10 million. In 1970 and 1971 they began to concentrate their efforts on dealer contracts and that source of business amounts to approximately 68% of their current volume. They are currently processing approximately 600 applications per month with a staff of five loan officers and a clerical staff of 15 people.

The Decision-Makers:

The decision-makers who acted as the principle judges in this research were the five installment loan officers of Lamar Savings. Two other experienced loan officers from other lending institutions in Austin also participated in the effort to assess the transferability of the policy models. A short background sketch of each judge is given below:

1) Judge #1 is an Assistant Vice-President and is the number two man in the installment loan department. He has nine years experience in installment lending and has been with this department for 3.5 years.

2) Judge #2 is a Vice-President and is the manager of the installment loan department. He has 14 years experience and has been with the department for three years. He has had previous experience with numerical credit scoring systems.

3) Judge #3 is an Assistant Vice-President and is third senior man in the department. He has seven years experience, most of which has been under the supervision of Judge #2 and concentrated in the area of collection of delinquent accounts. He has been in the department two years.

4) Judge #4 is a loan officer and has three years experience, all in this department.

5) Judge #5 is a loan officer and has two years experience, all with this department.

6) Judge #6 was recently the manager of a competitive lending institution in Austin and was hired into the department near the end of the project. He has had experience with dealer contracts and numerical credit scoring and a total of four years experience.

7) Judge #7 is the manager of a local credit union and has 14 years experience. His organization does not have any dealer contracts and they do not currently use numerical scoring techniques.

Two Functional Objectives of the Policy Models, Tentative Approvals and Surrogate Judgments:

In reference to the decision process being considered, the objectives of the modeling effort are to provide decision models that can be used both for screening the loan applications on a "tentative approval" basis and also pre-scoring the applications on a "surrogate judgment" basis. The difference between these two functions lies in the availability/non-availability of the credit rating for use as a predictor.

As depicted in Figure 5-1, the availability of a credit rating generally requires an input from the local credit bureau. At present, this input is obtained by telephone and is only available during normal working hours from Monday through Friday. If a merchant operates on the basis of six or seven days a week, this repreeents a significant impediment in his ability to sell and obtain timely financing for his merchandise. No action can be taken on a loan application taken after 4 p.m. Friday until at least 8 a.m. the

following Monday and the merchant must either delay delivery of the merchandise or assume the risk of possibly having to retrieve it if the loan is denied. This delay represents a tangible economic consideration to the dealer which he would like to minimize, especially in the case of those customers who appear to qualify for a loan on the basis of the other information available on the application form.

Whereas the credit rating is the most important single variable considered in the loan officers' decisions, they are generally unwilling to make a final commitment without it. However, it has been found that people with good credit ratings also generally score well on the other predictors. The problem is to determine if a policy, based on only the other predictors, can be defined that will identify a sufficient number of good accounts to justify its use without involving undue risk. If such a policy can be found, the credit executives would be willing to assume some increased risk in order to provide better service to the merchant. In this regard, a "tentative approval" would be an approval given in the absence of credit rating information which might require some modification of the contract such as "full recourse" if the credit rating was later determined to be unacceptable.

The pre-scoring or "surrogate judgment" application of the policy models would be accomplished with the use of the credit rating. In this case, the policy model would not suffer from the obvious structural deficiency of lacking the most important predictor. The rating from this model would be the expected rating that any application with that particular set of predictors would normally receive from a judge. Again, this could be used in two modes, if the judge finds no extenuating circumstances in the application, he can base his decision on a fixed threshold value and the predicted score; if he chooses to

consider other factors that are not accounted for in the policy model, he can use the predicted score as a starting point and base his decision on this score plus his evaluation of the extraneous factors.

Chronology of the Project:

The original impetus for this research came from a project in an advanced statistical applications course. Potential applications for policy capturing were being assessed and an attempt was made to model Judge #2's policy relative to a set of fictitious loan applications. This led to the discovery of four aspects of the problem that presented a challenge to defining viable policy models for loan officers' decisions. These were:

1) The methodology of ranking decision cases as was the standard procedure in most past Policy Capturing studies was untenable in this situation.

2) The generation of fictitious data cases in a designed factorial experiment led to many ridiculous combinations of predictors and detracted from the credibility of the process.

3) The problem of scaling and ranking of the predictors for use in standard statistical routines was not easily solved.

4) The modeling of an individual loan officer's decisions as a first-order function of the individual predictors neither produced satisfactory results nor was it intuitively satisfying to the loan officers involved.

The results of that initial effort did intrigue the participants enough to embark upon the research reported herein. The major activities of the project are now described.

Selection and Coding of Predictors:

The initial selection of predictors was suggested by the application form itself. The three senior loan officers (Judges 1, 2, 3) were each interviewed and asked how they defined and quantified the data on each of the input blanks. This resulted in the definitions and codes shown in Exhibit 2, Appendix B.

Retrieval of Data:

The category of loans under study consists of dealer paper on durable appliances. The original data base was obtained by randomly selecting 404 loan applications that had been processed by the loan officers during the period from May to September, 1971. Loan applications from two dealers were drawn from the files and coded according to the categories determined in the previous step. Those predictors such as age, income and equity which were directly transcribable were coded by the investigator. Those predictors that required some judgment in themselves were referred to the loan officer who had originally processed the application and he performed the coding.[3]

Analysis and Ordering of Predictor Categories:

Initial analyses were performed on a set of 404 coded

---

[3] In this respect the exercise of expert judgment in the measurement function as advocated by Sawyer (1966) and Einhorn (1972), (See Chapter II) was inherently accomplished as part of the procedure.

applications. For those variables that were naturally monotonic such as age and income, AID analyses were used to determine at which break-points, category definitions were meaningful. For those variables which were not naturally ordered, the categories were assigned ranks based upon the average approval rate for all applications in that category. This procedure is discussed in sections 3.2.4 and 3.2.6.2 of Anderberg (1971). At this point no assumptions were made as to the interval nature of the data and the predictor values were considered as ordered categories only.

Initial Analysis Efforts:

The categorized data from the sample applications was used in an initial series of computer runs that were directed primarily at answering the following four questions:

1) Can predictor variables and decisions from historical files be used to capture the judge's policy without requiring him to make new decisions on different cases?

2) Can the historical data taken from the file be used to build a model that adequately predicts the actual approvals/denials without considering Credit Rating?

3) What is the structure of the decision process and can the AID-Tree adequately depict this structure?

4) Can the AID-Program be used to detect, and isolate inter-judge policy differences?

With respect to question two, the reason for the exclusion of Credit Rating lies in the fact that Credit Rating is the only predictor that is

not available immediately after completion of the application and requires an input from a source external to the dealer-lender subsystem. (See Figure 5-1.)

Definition of a New Predictor:

A major result of these initial efforts was the classification and delineation of the concept of Financial Reference. Originally this predictor had been stated as Credit Habit and was given the subjective categories of good and bad. The initial analysis indicated that it was the second most important predictor. Its importance led to a discussion of what constituted a good or bad Credit Habit in the opinion of the senior loan officers. After some debate, these officers agreed that the concept would be better labeled as Financial References and was a measure of the depth and quality of the applicant's borrowing habits. The specific categories they defined were:

Major--the credit references listed by the applicant include credit depth on experience with major lending and financial institutions which do not take undue credit risks and from whom a reliable credit reference might be obtained. This category includes banks, savings and loans associations, major chain department stores, and national based lending agencies.

Minor--the credit references listed by the applicant include only local stores and usinesses from whom a direct credit check was not readily available or necessarily reliable. The credit references do not indicate significant depth of credit, i.e., experience of the applicant in managing his finances to meet a continuing repayment of a loan of size comparable to that being applied for. Examples of applicants in this category would be those listing only small revolving

charge accounts, such as oil companies or those listing only local concerns which carry their own short-term credit.

Other--the credit references listed by the applicant indicated reliance upon lending institutions specializing in high-risk high-interest loans. The references indicated above average dependence upon borrowed money or lack of capability to manage one's money effectively to meet normal living expenses. Those people listing no references at all were included in this category.

Refinement of the Sample Data Set:

A second refinement of the data set involved the concentration of effort on modeling the policy of the judges toward only one of the dealers, namely, Dealer A. This resulted in the reduction of the data set to 224 cases, of which 100 had been approved and 124 had been denied. These data are shown in Exhibit 4, Appendix B.

Generation of Codified-Predictor Data and Collection of Judges' Decisions:

Using the data entries for each of the 224 cases (See Exhibit 4.), codified predictor data for each application was generated in terms of the descriptors for each of the predictor categories. The predictor descriptors for each of the cases were displayed on a computer printout (Exhibit 5, Appendix B) and used by the judges to make new decisions for each application. The judges were given one set of codified predictor data which included all of the predictors except Credit Rating and were asked to make their decision in the context of "tentative approval". Later, they were given a second set of data which included the Credit Rating and were asked to make the decision in the context of granting final "approval" or "denial".

As part of this activity, the two junior loan officers, #4 and #5, were brought into the project. They were informed of what had previously transpired and the objectives of the experiment; each of the three senior loan officers discussed his policy as depicted by the AID-Tree. The discussion that followed was possibly one of the more enlightening aspects of the entire project. The major points of this discussion revealed the following situation:

1) Since the installment department was in a state of rapid expansion, new people were hired and put to work immediately without benefit of any formal training program. Loan officers were generally assigned to collection activities initially and later allowed to evaluate applications under the supervision of the senior loan officers. However, in this latter phase the supervision generally took the form of getting a senior loan officer's approval on the final decision about an application. Although this procedure does provide some feedback in the cases where the junior and senior loan officers reach a different decision, it apparently does not suffice to provide the junior loan officers with a good understanding of how the senior loan officers evaluated each of the entries on the application form. Thus, the discussion of the applications to be rated appeared to provide the catalyst for the first comprehensive exchange of opinion among all five of these loan officers.

2) The concept of Financial Reference as previously agreed upon by the three senior loan officers was not a part of the policy of the junior loan officers.

<u>Generation of Policy Models</u>:

Upon completion of the evaluations by each of the loan officers, policy models for each of the officers were developed using

th· techniques discussed in Chapter IV. These models are presented and discussed in Chapter VI.

At this point a different problem became apparent in using these models as a predictive tool. The problem is that of which loan officer's model to use on which application. In the current operating scheme, the incoming applications generally are not routed to any particular individual, but are acted upon by whomever is available. This factor not only posed a problem for adequately evaluating the performance of a given loan officer's model but also has the effect of making an applicant's chances for getting a loan a function of which judge evaluates it, and what his idiosyncrasies may be. Therefore, several voting strategies were investigated. The strategy that best matched the actual decisions was determined to be one in which a loan was approved if four of the five of the individual loan officers approved it.

Validation by Field-Test:

Two of the original goals of this research were to analyze the problems that might be encountered in the implementation of these policy models in the real environment and to demonstrate the adequacy of the models by cross-validation. To this end, the appliance dealer who was the source of the 224 applications that were used in the modeling phase was approached and asked to participate in a one-month field-test. This field-test took the form of having the dealer's clerical staff fill out the codified application form, Exhibit 6, Appendix B, in addition to the normal application form on each sale that required financing during July, 1972. In order to accomplish this task, the clerical staff was briefed on the project and given a set of instructions

as shown in Exhibit 6, Appendix B. Upon receipt of the original application, the loan officers also filled out the coded form but made their decision in the same manner they had always done. During the month, 150 usable applications were collected which were then analyzed in several ways. These analyses included:

1) evaluation of model sensitivity coding differences between "expert measurement" of the loan officers and the "non-expert measurement" of the clerical staff,
2) cross-validation of the previously derived individual and voting policy models,
3) evaluation of the average time and range of times the applications were in process.

An Insight into Transferability:

One of the interesting findings arising out of the review of past experience with numerical credit scoring systems is the fact that such systems have been "borrowed" as opposed to "tailored" for new enviornments. A natural question then is "are policy models captured in one environment amenable to being transferred to new environments?" Although an adequate answer to this question would require a major research project itself, an attempt was made to get an indication of the potential transferrability of the models. In this effort two experienced loan officers who were managers of other lending agencies in Austin were interviewed and asked to make decisions based on the codified predictor on the 224 applications. One of these individuals, Judge #6 had recently been hired by the participating installment credit department but had not yet worked with them long enough for their

policies to influence his opinion.

Summary:

This chapter has briefly reviewed the past work in numerical credit scoring and has described the environment in which this experiment was conducted.

A short chronological description of the research activities relative to capturing and evaluating the policy models for the loan officers has been given. The various forms and instructions generated and used in these activities have been included in Appendix B. The results, conclusion and implications of this project are discussed in Chapter VI.

CHAPTER VI: ANALYSIS, RESULTS, AND IMPLICATIONS OF LOAN OFFICER POLICY MODELS

Introduction:

In documenting the results of this research, it might be sufficient to simply list the final set of equations and their cross-validation hit rates. However, many initial and intermediate results were obtained, each of which influenced the direction of the research activities that followed to some extent. Not all of the results were positive, but even those that were negative added to the knowledge of the investigator. It would seem somewhat pretentious not to discuss all of these results for two reasons; a) they may answer questions of later investigators and help them avoid unnecessary steps, or b) they may provide the impetus for better ideas or decisions than those which were pursued by this investigator. Therefore, this chapter includes the intermediate results as well as the final equations in the hope of providing a better insight into what worked and what did not work. The breadth of the results presented in this chapter may create a problem in assimilating the information presented. To alleviate this problem, the presentation has been further segmented into topical sections and the following directory is provided.

Section I.

Analysis and Refinement of the Original Sample:

The original sample of data included 404 applications which were randomly selected from the files of the lending institution. Of these, 270 applications had been submitted by a high volume dealer (Dealer A) who had been participating in a dealer contract arrangement with the lender for a period of approximately six months. The other 134 applications were from a lower volume dealer (Dealer B) who had been working on a dealer contract basis with the lending institution for several years.

The data on each application was coded in terms of one of the category descriptors shown in Exhibit 2, Appendix B. For those predictors that were not naturally monotonic, the categories were rank ordered according to the average success rates for all applications within each category. The data in Table 6-1 reflects the success, or approval, rate for each category of each predictor and was the basis for the category rankings.

TABLE 6-1

BY-CATEGORY SUCCESS RATES USED FOR RANKING OF PREDICTOR CATEGORIES

Variable \ Category	1	2	3	4	5	6	7
Age	.50	.57	.58	.79	.89		
Marital	.59	.63	.63				
Local Family	.58	.61					
Draft	.00	.60					
Telephone	.35	.63					
Average Time	.53	.62	.68				
Residence Cat.	.25	.51	.72	.88			
Employ. Type	.48	.62	.64	.65	.72		

Table 6-1 (cont.)

Variable \ Category	1	2	3	4	5	6	7
Income	.44	.49	.64	.48	.59	.74	.74
Bank	.33	.67					
Credit Habit	.12	.70					
Loan Amount	.53	.59	.56	.66	.60	.70	
Loan Term	.68	.53	.55	.57			
Necessity	.54	.69					
Equity	.56	.63	.74				
Credit Rating	.10	.55	.65	.94			

The initial analysis of the data set using both bi-variate contingency tables generated with Anderberg's (1971) GCORR program and output from the AID4 program led to several refinements and modifications to the predictor set. These included:

1) Elimination of draft as a variable since draft eligibility automatically eliminated an applicant from consideration based on company policy.

2) Elimination of debt as a predictor since useable data was available on less than 1/3 of the applications.

3) Reduction in the number of categories required to describe the loan amount and income predictors.

4) Inclusion of self-employed and military officers into the executive category based on indifference of initial AID analyses to their classification in separate categories.

5) Elimination of those applicants achieving loans through "full recourse" guarantees of the dealer.

These modifications resulted in a refined data set of 389 applications, 262 from Dealer A, and 127 from Dealer B.

Initial Analysis Disregarding Credit Rating as a Predictor:

The 389 data cases were used in exploratory efforts to build a model that would adequately predict the actual response without the use of credit rating as a predictor. Various modeling alternatives including multiple regression, multiple classification analysis, and AID-Trees were pursued.

These efforts were generally unsuccessful, resulting in first-order and second-order equations that could account for only 20-25% of the variability and AID-Trees that required 61 splits (62 mutually exclusive groups) to account for 60% of the variability.[1] The primary reason for the poor quality of these early models was assessed to be structural deficiency in the models. The actual decisions had been made considering the Credit Rating of the applicant. Attempts to model these actual decisions without using Credit Rating as a predictor represented omission of the single most important variable in the judgment process.

A second task of the preliminary analysis was to determine if a better fitting model (in terms of unexplained variability) might be obtained by using a continuous criterion. To this end, the loan officers had been asked to rate each of the applications on a scale

---

[1] Throughout the initial analyses, discussed in Section I, Chapter VI, the square of the multiple correlation coefficient (R^2) was used as the criterion of goodness for making comparisons and decisions. Later the hit rate (see Chapter IV, page 84) was adapted as a more appropriate criterion and used for the remainder of the research.

of 1 to 5. Figure 6-1 shows the comparison of the variability explained as a function of the number of mutually exclusive groups for both the binary and scaled criteria. Only Judge 1 and Judge 3 had processed a sufficient number of the original applications to perform the analysis, but in both cases their scaled criterion values were less predictable than their binary decisions. Review of the residuals for each of these approaches revealed that most of the added variance experienced in the scaled rating models was attributable to those applications that were denied, possibly indicating a lower thresholding effect in the judges' policies, i. e., a good application had measurable levels of "goodness" but a bad application was just "bad". These results would suggest that having the judge rate or rank the data cases in a manner different from their natural decision process would tend to produce erroneous policies models.

Another exploratory approach involved attempting to define "factors" entering into the judgment. In this effort Factor Analysis computer routines were used to combine 13 predictors into five "factors". These five factors were capable of explaining some 60% of the internal variability within the predictor set. However, the resulting combinations of variables in each of the factors were not particularly meaningful to the loan officers who stated that they really thought of loan applications in terms of the following four factors:

1) Capacity: capability to service debt
2) Character: stability of applicant
3) Credit: past credit record
4) Condition: particulars of loan and risk to lender.

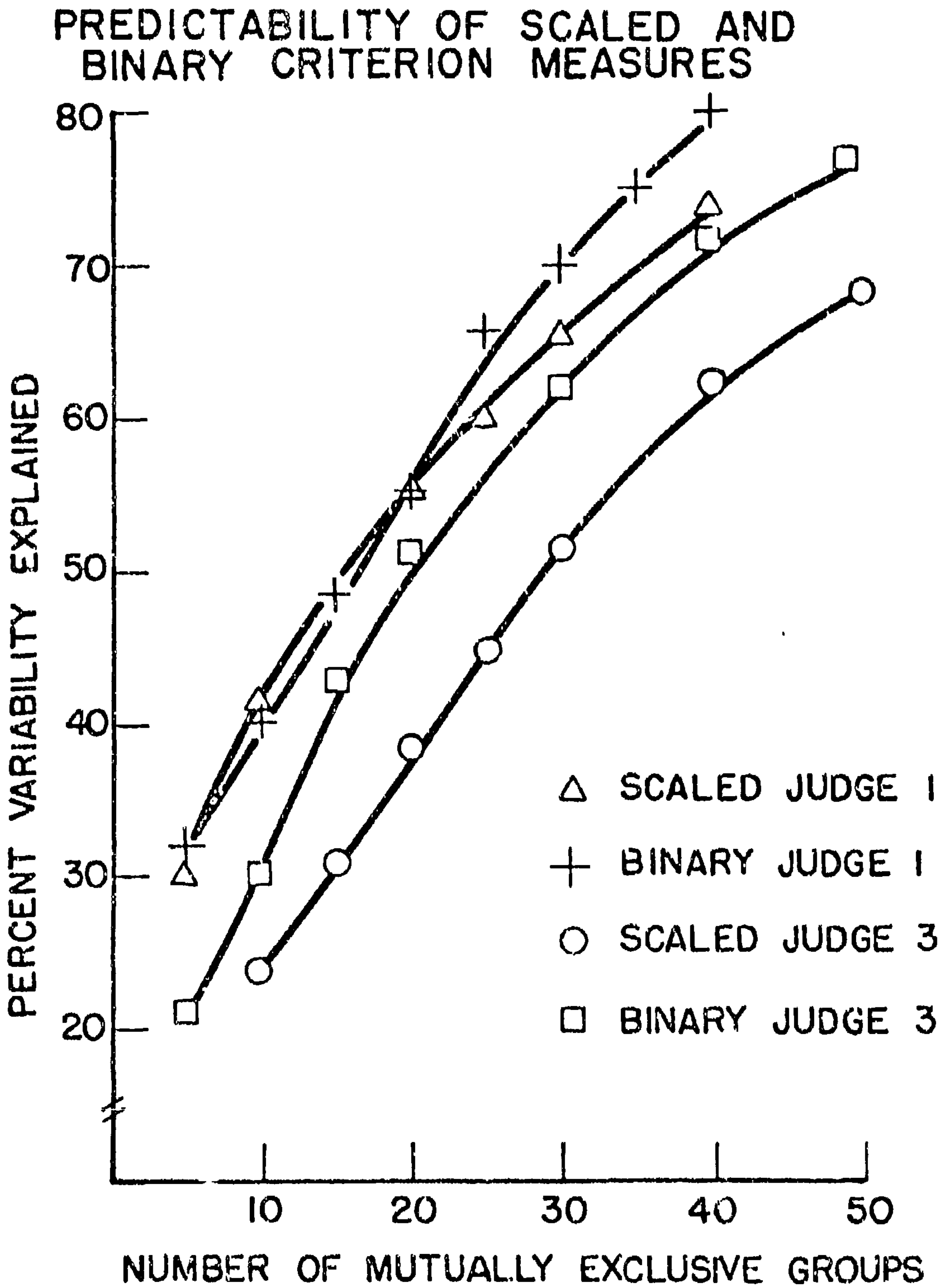

PREDICTABILITY OF SCALED AND BINARY CRITERION MEASURES
80
70
60
50
40
30
20
PERCENT VARIABILITY EXPLAINED
10
20
30
40
50
NUMBER OF MUTUALLY EXCLUSIVE GROUPS
△ SCALED JUDGE 1
+ BINARY JUDGE 1
○ SCALED JUDGE 3
□ BINARY JUDGE 3

FIGURE 6-1

Since one of the underlying objectives of the research project was to model known policies of the loan officers in meaningful terms, the "factors" that were mathematically defined were discarded and not considered further.

The analysis and comparison of the AID-Trees for each of the dealers, on an individual basis, did result in illumination of an important point. The trees showed an apparent difference in the policy of the lending institution toward each of the dealers. In the case of Dealer A, only 52% of the applications were approved while 78% of the applications from Dealer B were approved. Further, AID-Trees reflected that different variables were important in explaining the variability in each of the policies. For Dealer A, the most important predictors, as indicated by the first few splits of the AID-Tree, include Bank Account, Age, Equity, and Income. For Dealer B, the most important predictors were Residence, Employment Type, Loan Term, and Local Family. This apparent difference was discussed with the judges and they reflected that there actually was a difference in their attitude toward the different dealers. They noted that this difference was not explicit but was implicit and arose from their working relationship with the individual dealers and the types of applications they received from each. Specific causes of the differences were:

1) Dealer B has a more affluent and mobile clientele than did Dealer A.

2) They had been operating with Dealer B for a much longer time and they expected that some pre-screening of the applications, based on their accept/reject experience, occurred before the applications were forwarded.

3) Dealer A was more helpful in recovery and disposal of repossessed merchandise and so they were relatively more interested in the variables that reflected the applicant's "capability" to pay than they were in those reflecting his "credit" or "character".

Preliminary analysis for each of the three senior loan officers did indicate differences in the variables each considered the most important and served to convince them that the AID-Trees were a viable means of portraying and analyzing their policies. For example, Figure 6-2 reflects the success rates for each of the categories of the predictor Marital Status for the three senior loan officers. It indicates that being single is somewhat of a detriment in the eyes of Judge 1 while being divorced is a detriment from Judge 3's viewpoint. It is interesting to note that during the formulation of the predictor categories, all three judges had expressed the opinion that a married applicant had a better chance of getting a loan, but their responses reflected this to be true only in the case of Judge 1. However, a cautionary note is in order here! The assertion of a given attribute being "better" should be interpreted in the contexts of "all other things being equal". Hence, the zero-order correlation between "married" and "approval" may not reflect the true importance of the marital status predictor in the situation where there are other intervening predictors that are not independent of marital status. Nevertheless, the use of plots such as Figure 6-2 was found to be a good catalyst for discussion of the individual loan officer's policies.

Although differences between judges and dealers were detectable in the early analysis, the inability to use only historical

SUCCESS RATE VS. MARITAL STATUS

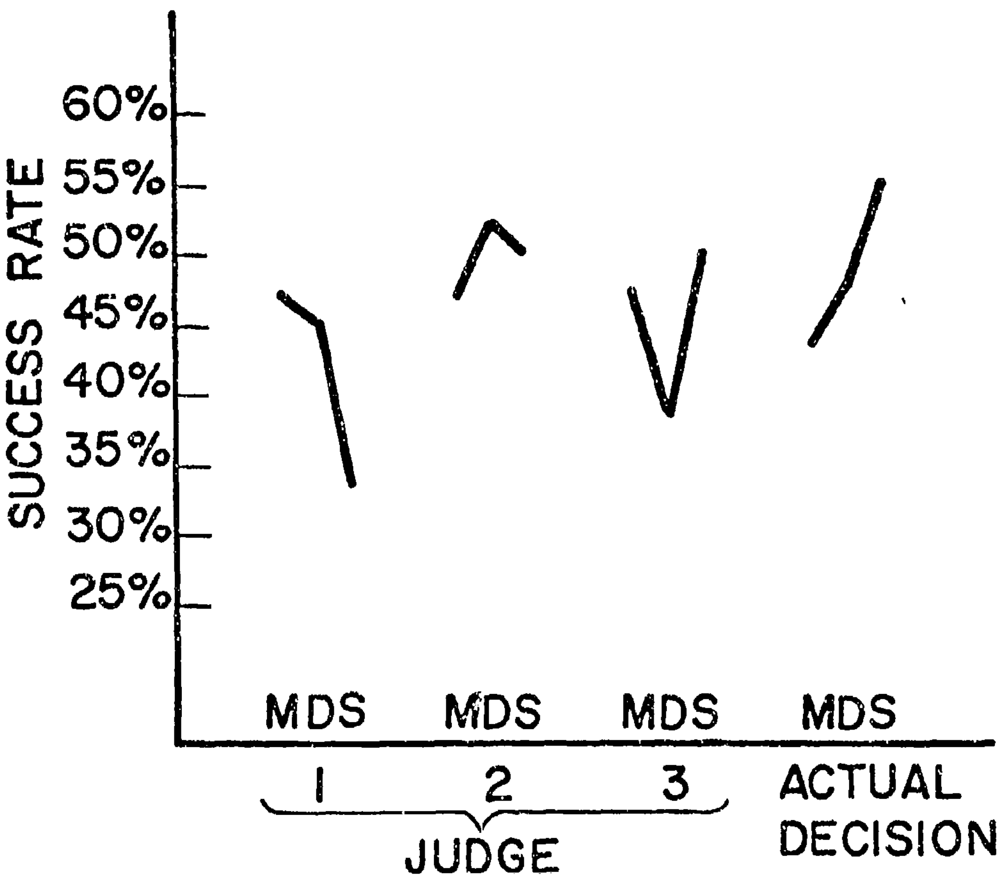


FIGURE 6-2

decisions to obtain models that were sufficiently predictive caused considerable consternation and re-analysis of the decision process and modeling methodology. Three basic problems were identified:

1) The transformation from an AID-Tree into a configural effects regression model requires a different conceptualization of interaction among the predictors than is indicated by the standard interaction term of the form $b_{12}x_1x_2$.

2) Decision processes probably cannot be modeled very successfully if all of the important variables that entered into the original decisions are not included in the model. For the loan officers, there is no way to partial out the effect of Credit Rating from the actual responses that had been made with it being considered.

3) In situations where the number of data points is relatively small compared to the number of predictor categories (389/53 in this data), uncontrolled variation in the input data used by the judges must be minimized to improve the "signal" to "noise" relationship.

The solution to these three problems was pursued in two directions, the first being the methodological development as discussed previously in Chapter IV; the second being the generation of the coded predictor data and the re-evaluation of this data by the three original senior loan officers and two junior loan officers (Judge #4 and Judge #5).

The re-evaluation of the applications based on the predictor "descriptors" was undertaken for two purposes:

1) to eliminate from consideration any extraneous

information that had appeared on the original application and thereby assess the effect of controlling the information on the application form, and,

2) to facilitate interjudge comparisons by allowing comparison of each loan officer's opinion on each loan.

This procedure reflects adherence to Hoffman's (1960) principle of experimental control. Although the raw historical data did not permit definition of adequate policy models, it did provide a judgment data set reflective of the population of actual decision cases that are encountered in the real world.

Prior to regeneration of the predictor data, further refinement of the predictors was accomplished with the final set of predictors consisting of the 16 variables shown in Exhibit 3, Appendix B. The refinement was accomplished on the applications from Dealer A. Of the original 262 applications, only 224 could be retrieved from the installment company's files due to the fact that the files had been relocated and reorganized and some accounts had been closed out during the period of the initial analysis. This situation in itself reflected one of the difficulties of performing a Policy Capturing analysis with the use of actual data from a functioning organization. In order to cause a minimal disturbance on the on-going system, the original applications had been left in the system and were not totally accessible at a later date.

Section II.

Decisions Based on the Codified Data Set: (Without Credit Rating)

Each of the judges was given a set of the Codified Predictor Sheets with all predic ors except Credit Rating and asked to either grant "tentative approval" or "reserve judgment". During this process, considerable discussion resulted relative to which predictors were important and which were not important. This was one of the particularly interesting and beneficial aspects of the study in that it represented the first time all of the five loan officers had extensively discussed their philosophies and policies relative to evaluating loans.

Since the objective of the decisions made without benefit of the Credit Rating was to identify only the best prospects for "tentative approval", the loan officers defined seven conditions, any one of which would automatically prohibit tentative approval regardless of the overall quality of the application. These conditions were:

1) under 21 years of age
2) under 2 years residence time and under 2 years job time
3) over 5 years residence time and still renting
4) over 25 years old without major credit references
5) under $400 income
6) no bank account
7) "other" financial reference (code 1)

These conditions reflect a conservative bias on the part of the judges since they were viewed as significant weaknesses in the

applications that made a positive decision, in the absence of a credit rating, unduly risky. These rules were implemented in the models by defining a binary predictor that indicated presence of one or more of the seven conditions. For want of a better term, this predictor was labeled "Dings". Use of Dings as a predictor allowed much better fitting models for each judge's "without credit rating decisions". However, the match between the decisions predicted by the models and the actual approvals/denials was still not as good as desired.

There was considerable variation in which loans the judges individually gave "tentative approval" and only 66% of the 224 decisions were unanimous. More importantly, only 20% of the applications received unanimous "reserve-judgment-without-credit" decisions. This is partly attributable to denying the judges use of their primary predictor and partly due to the fact that, in making the "tentative approval" decision, they were really making a different decision than they were normally accustomed to making. In an effort to determine if a composite evaluation might be better for isolating just the very best prospects, various voting strategies were investigated. It was found that a strategy based on granting tentative approval if four of the five judges had individually given "tentative approval" to the application and did the best job of identifying the top 25-30% of the applications while minimizing the inclusion of applications which had actually been denied. The criterion of 25-30% was arrived at based on discussion with loan officers relative to what percentage of "tentative approvals" would be required to make a policy model beneficial to them. Voting models considering the "Dings" reflected some over-conservatism in that only 71 out of 224 cases would have received tentative approval and 11 of these 71 had actually been denied. Voting models that did not use "Dings" as a

predictor performed approximately the same since 74 of the 224 cases would have been tentatively approved with 12 mistakes. The regression coefficients for these models are presented in Table 6-2.

Again, one of the primary results of this effort came from discussion and clarification, in the judges' minds, of which attributes were indicative of abnormal risk.

Decisions Based on Codified Data Set: (With Credit Rating)

Each judge was given Codified Predictor Sheets which included the Credit Rating for each of the applicants. They were asked to make their decisions as if they were final approvals or denials on the loan. This set of decisions provided the data for comparison of the five judges on an inter-judge basis. Decisions on the Codified Predictor Sheets were made twice by each judge, thus providing information on the reliability (percent of decisions that were the same on both trials) for each judge. Comparison of the decisions made on each set of codified predictor data with the actual decisions, that had been made previously from the original application form, provided insight into the amount of "information" lost in the process of codifying the predictors. Table 6-3 contains the hit-table for the coded versus actual decisions and the reliability of each of the judges. Since each of the original applications had been evaluated by only one of the three senior judges, the matches between the original decisions of each judge and his decisions on the corresponding sets of codified predictors is also shown.

TABLE 6-2

TENTATIVE APPROVAL MODELS

$$\text{PREDICTED SCORE} = \text{CONSTANT} + \sum_i \text{COEFFICIENT}_i \times \text{PREDICTOR}_i$$

Predictor:		Coefficient
1.	Constant	-1.377
2.	Age	- .0125
3.	Marital	- .0542
4.	Local	.0355
5.	Phone	.2311
6.	Jobtime	.0536
7.	Restime	.0387
8.	Rescat	.0918
9.	Type Employment	.0230
10.	Income	.0164
11.	Bank Account	.1200
12.	Financial Reference	.3002
13.	Loan Amount	.0328
14.	Loan Term	- .0352
15.	Necessity	.0371
16.	Equity	.0275
17.	Dings	- .1200

Model without dings includes only first 16 terms.

TABLE 6-3

CODED vs ACTUAL DECISION STATISTICS
(C/A)

	HIT-TABLE										
Judge	No/ No		Yes/No		No / Yes		Yes/Yes		%Correct		Reliability
Trial(1, 2)	1	2	1	2	1	2	1	2	1	2	
1	106	112	18	11	19	25	81	75	84	84	(90%)
2	101	90	19	34	12	4	88	96	86	83	(89%)
3	102	100	22	24	18	11	82	89	82	84	(90%)
4	93	99	13	25	31	21	87	79	80	80	(80%)
5	100	110	24	14	13	24	87	76	84	84	(85%)
"4/5" VOTE	110	--	14	--	17	--	83	--	86	--	

(Based on Judge Making Original Decision)

Judge	Cased Decided	% Correct
1	112	86.5%
2	24	85.0%
3	88	87.5%

The key point derived from the comparison of the judges' decisions made on the codified predictors with their original decisions based on the full application is that codifying the data only slightly affected the decisions. This assertion is made since the reliability of each of the three senior judges was approximately 90 percent and the correspondence of their decisions based on the two forms of data ranged between 85 and 87.5 percent.

The inclusion of the Credit Rating predictor and the refinement of the categories of the other predictors resulted in attainment of a vastly improved fit to the data. Figure 6-3 shows the AID-Tree developed using the actual decisions as the criterion. In the total tree[2], nine splits occurred leaving ten mutually exclusive groups which accounted for 76 percent of the variability in the data.

Table 6-4 shows the ranked importance for the variables relative to their power to explain variability over the total predictor space and also over the predictor subspaces defined by the AID splitting process. The individual predictors are ranked according to their average BSS/TSST and BSS/TSS(i) values. The data in this table would indicate that CREDIT RATING, FINANCIAL REFERENCES, BANK ACCOUNT, LOAN TERM, TYPE EMPLOYMENT and PHONE are the most important variables in the decision process. Since they hold ranks 1 through 6 in both rankings, they are important for all applicants. MARITAL STATUS, LOCAL FAMILY, and NECESSITY hold very low ranks in both rankings and are the least important predictors.

---

[2] Only eight of the splits actually appear on Figure 6-3 due to limitations in this version of the tree printout. The expanded version of the AID-Tree shown in later examples can portray 12 levels of the splitting process.

AID-TREE FOR ACTUAL DECISIONS WITH CREDIT RATING

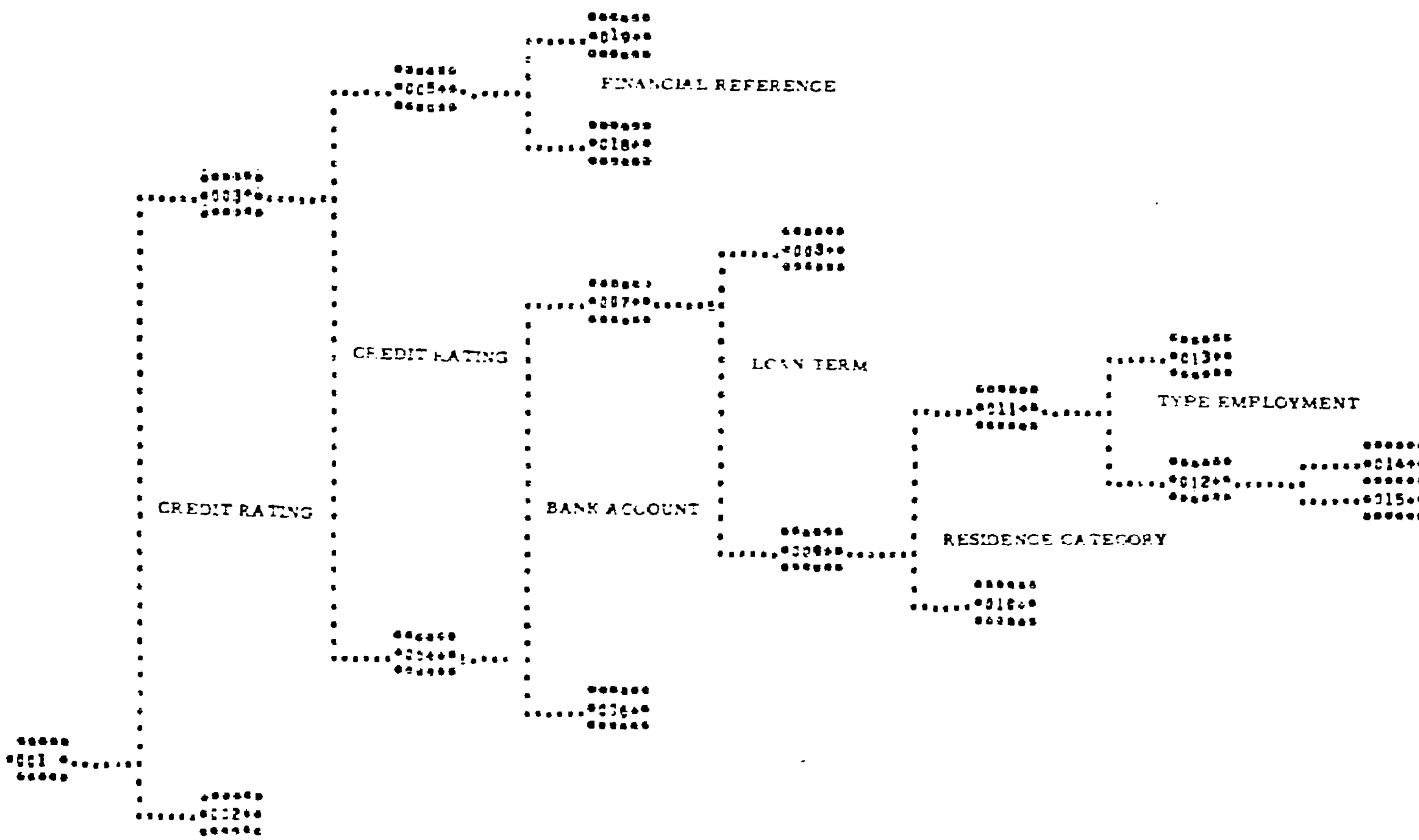


FIGURE 6-3

TABLE 6-4

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 12 GRAND MEAN = .161534 ROUND CRIT. = .010TSST = 55.4

		SUB(VARIANT) SUMBSS/TSST	RECONSTRUCTED SUMBSS/TSST	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	AGE	.07213	.07213	.00620	1.03229	.00004	44.65030	12	12.0
2	MARITL	.04772	.04772	.00463	1.16499	.00002	29.54225	12	13.0
3	LOCAL	.01220	.01220	.00135	1.32483	.00000	7.55339	12	16.0
4	PHONE	.14824	.14824	.01487	1.05099	.00022	104.15052	12	5.0
5	JOBTIM	.07599	.07689	.01227	1.91430	.00015	47.59743	12	11.0
6	RESTIM	.04230	.04230	.00564	1.60014	.00003	26.18393	12	14.0
7	RESCAT	.11115	.12126	.01359	1.34464	.00018	75.06468	11	7.0
8	TYPEMP	.13766	.13766	.01099	.95840	.00012	85.21855	12	6.0
9	INCOME	.08478	.08478	.00941	1.33236	.00009	52.48242	12	9.0
10	BANKAC	.25180	.45217	.05064	1.43934	.00256	261.34742	10	2.0
11	FINREF	.28126	.28126	.03699	1.57807	.00137	174.11977	12	3.0
12	LOANAM	.08728	.08728	.01059	1.45667	.00011	54.03219	12	8.0
13	LOANTM	.16920	.18458	.01479	.96129	.00022	114.26462	11	4.0
14	NECESS	.02635	.02635	.00259	1.17720	.00001	16.31315	12	15.0
15	EQUITY	.08243	.08243	.00577	.84053	.00003	51.02742	12	10.0
16	CRERAT	.61444	.73733	.12482	2.03145	.01558	456.45196	10	1.0

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 12 GRAND MEAN = .510625 ROUND CRIT. = .010TSST = 55.4

		SUB(VARIANT) SUMBSS/TSS(I)	RECONSTRUCTED SUMBSS/TSS(I)	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	AGE	.53464	.53464	.04805	1.07853	.00231	104.70327	12	7.0
2	MARITL	.23512	.23512	.02158	1.10151	.00047	46.04487	12	13.0
3	LOCAL	.12633	.12633	.01624	1.54223	.00026	24.74087	12	16.0
4	PHONE	.59696	.59696	.04183	.84080	.00175	116.90730	12	6.0
5	JOBTIM	.25778	.25778	.01694	.78871	.00029	59.48232	12	12.0
6	RESTIM	.16459	.16459	.01916	1.39695	.00037	32.23397	12	15.0
7	RESCAT	.33819	.36893	.02619	.85193	.00069	72.25087	11	11.0
8	TYPEMP	.61202	.61202	.05020	.98433	.00252	119.85710	12	5.0
9	INCOME	.37598	.37598	.01914	.61101	.00037	73.63162	12	10.0
10	BANKAC	.75496	.90595	.07378	.97730	.00544	177.42037	10	2.0
11	FINREF	.80190	.80190	.07155	1.07064	.00512	157.04289	12	4.0
12	LOANAM	.42555	.42555	.04663	1.31499	.00217	83.33971	12	8.0
13	LOANTM	.74687	.51695	.06405	.94074	.00410	159.94074	11	3.0
14	NECESS	.21034	.21034	.02320	1.32366	.00054	41.19260	12	14.0
15	EQUITY	.40444	.40444	.02783	.82559	.00077	79.20565	12	9.0
16	CRERAT	1.11047	1.33251	.12419	1.11844	.01542	260.95583	10	1.0

RANKED IMPORTANCE OF PREDICTORS

Section III.

The Individual Policy Models:

Using the responses to the Codified Predictor Sets as the criterion, first-order, second-order, and AID-derived "local" models were developed for each[3] of the five judges and for the four of five voting strategy.

The First Order Models:

The coefficients for first-order regression models of the form

$$\text{Predicted Score} = \text{Constant} + \sum_i \text{coefficient}_i \times \text{predictor}_i$$

were obtained following the iterative model determination/cross validation procedure described in Chapter IV. Implicit in these models is the assumption that the predictor categories are ordered according to the codes given in Exhibit 3, Appendix B and that a constant change in the Predicted Score will result from a unit change in the level of a Predictor Variable. This assumption was made in an attempt to avoid idiosyncratic results that could result from fitting the 224 data points with a binary regression model which had 53 predictors. No attempt was made to test individual terms for statistical significance. However, it is apparent that any variable whose coefficient is smaller in magnitude

---

[3] Second-order models were only derived for Judge 1, 2 and 3.

than .002 would have an effect on the total score of less than .01 (since the range of values for the variables is between 1 and 5) and should probably be ignored. The first order coefficients for the 5 judges are shown in Table 6-5. The best cut score reflects that point where the maximum hit rate occurs.

The Second-Order Models:

Second-order regression models of the form

$$\text{Predicted Score} = \text{Constant} + \sum_i \text{coefficient}_i \times \text{predictor}_i$$

$$+ \sum_i \sum_j \text{coefficient}_{ij} \times \text{predictor}_i \times \text{predictor}_j$$

were also analyzed. In the absence of substantive theory from which to hypothesize the terms in the second-order models, they were hypothesized by generating all possible first-order and second-order terms and using a stepwise regression routine to eliminate those terms not statistically significant at the .01 probability level. Due to computational limitations, the identification of potential terms was done incrementally with 16 first-order and approximately 45 second-order terms being considered in each of three runs. All terms identified by this procedure were included in the second-order model regardless of their "meaningfulness" or "interpretability". The iterative procedure was then used to determine the coefficients in Table 6-6 for the hypothesized model for each judge.

AID-Split Models:

AID-Split models of the form

TABLE 6-5

FIRST-ORDER MODEL COEFFICIENTS

Predictors: Judge	#1	#2	#3	#4	#5
Constant:	-1.1178	-.7174	-.7337	-.6865	-.9802
Age:	- .0198	.0017	-.0183	-.0181	-.0298
Marital:	- .0672	-.0051	-.0070	-.0219	-.0485
Local:	.0810	.0057	.0229	.0773	.1168
Phone:	.1595	.1208	.0570	.1178	.0901
Jobtime:	.0261	.0267	.0219	.0476	.0816
Restime:	- .0362	-.0051	-.0129	-.0449	.0068
ResCat:	.0685	.0453	.0161	.0427	.0563
Type Employment:	.0327	-.-135	.0019	.0261	.0184
Income:	- .0473	-.0208	-.0097	-.0336	-.0048
Bank Acct.:	.0506	.0753	.0673	.0200	.0578
Financial Ref.:	.2033	.0794	.1086	.1102	.1645
Loan Amt.:	- .0002	-.0250	.0039	-.0036	.0087
Loan Term:	- .0403	-.0470	.0226	-.0601	-.0733
Necessity:	.0595	-.0526	.0052	.0387	.0104
Equity:	- .0013	.0121	.0000	.0330	-.133
Credit Rating:	.2400	.2950	.2873	.2820	.2274
Best Cut Score:	.49-.50	.47	.50-.52	.37-.41	.40-.41
Hit Rate:	.91	.92	.91	.88	.90

TABLE 6-6

SECOND-ORDER MODEL COEFFICIENTS

Judge #1		Judge #2		Judge #3	
Constant	= -.4833	Constant	-.5496	Constant	-.4676
Phone	= .1438	Phone	.0918	Credit Rating	.2872
ResCat	= -.0071	Bank Account	.0000	Fin. Ref.• Bank	.0288
Bank Account	= .0068	Financial Reference	.0139	Fin. Ref.• Phone	.0144
Financial Reference	= .0000	Loan Term	-.0203	ResCat	-.0053
Credit Rating	= -.0098	Credit Rating	.2890		
Phone• Credit	= .0172	Marital• Jobtime	.0226		
Phone• Income	= -.0165	Marital• Necessity	-.0215		
Phone• Jobtime	= -.0012	Phone• Bank	.0105		
Local• ResCat	= .0259	Bank• Fin. Ref.	.0247		
Marital• Loan Term	= -.0259	Loan Term• Necessity	-.0213		
ResCat• Type Emp.	= .0063				
Bank Acct.• Necessity	= .0232				
Fin. Ref.• Credit Rtg.	= .0896				
(Bank Acct.)^2	= -.0004				
Best Cut Score:	.43		.52		.47-.48
Hit Rate:	.92		.92		.90

$$\text{Predicted Score} = \sum_i \text{group mean}_i \times GM_i$$

$$\text{where: } GM_i \begin{cases} = 1 & \text{if applicant is in group } i \\ = 0 & \text{otherwise} \end{cases}$$

were obtained by analyzing the application data with the AID4UT program. The models as shown in Figures 6-4, through 6-8, represent the first 5 or 6 splits of the tree, at which point subspaces (Figures 6-9 through 6-13) were defined and local models were developed. In Figures 6-4 through 6-8, the group means are denoted by "A_i" and the split variable is denoted by "$X_j^{(k,l)}$". The subscript "j" represents the identification number of the split predictor per Exhibit 3, Appendix B. The superscripts "(k, l)" denotes the categories of the split predictor that are in each group.

<u>Local Models:</u>

Local models of the form

$$\text{Predicted Score} = \text{Constant} + \sum_i GM_i \times (\text{Local Model}_i)$$

$$\text{where: } GM_i \begin{cases} = 1 & \text{if applicant is in subspace } i \\ = 0 & \text{otherwise} \end{cases}$$

Local Model i = First or second-order model pertaining only to those applicants in subspace i.

were derived based on analysis of the AID-Trees. In the process of developing the local models, several alternative formulations within the various subspaces were investigated. The general result was that,

AID SPLIT DIAGRAM
JUDGE 1

1. $Y = A_1 U + E^{(1)}$

$A_1 = .44$, $N_1 = 224$

2. $Y = A_3 X_{16}^{(2,3,4)} + A_2 X_{16}^{(1)} + E^{(2)}$

$A_3 = .66$ $N_3 = 148$

$A_2 = .01$ $N_2 = 76$

3. $Y = A_5 X_{11}^{(2,3)} + A_4 X_{11}^{(1)} + A_2 X_{16}^{(1)} + E^{(3)}$

$A_5 = .76$ $N_5 = 129$

$A_4 = .00$ $N_4 = 19$

4. $Y = A_7 X_{10}^{(2,4,3)} + A_6 X_{10}^{(1)} + A_4 X_{11}^{(1)} + A_2 X_{16}^{(1)} + E^{(4)}$

$A_7 = .88$ $N_7 = 99$

$A_6 = .37$ $N_6 = 30$

5. $Y = \underbrace{A_8 X_{16}^{(3,2)}}_{\text{SS-1}} + \underbrace{A_9 X_{16}^{(4)}}_{\text{SS-2}} + \underbrace{A_6 X_{10}^{(1)}}_{\text{SS-3}} + \underbrace{A_4 X_{11}^{(1)} + A_2 X_{16}^{(1)}}_{\text{SS-4*}} + E^{(5)}$

$A_8 = .76$ $N_8 = 46$

$A_9 = .98$ $N_9 = 53$

*For Judge 1, group 2 and 4 were combined into a common subspace due to the homogeneity of their mean criterion values.

FIGURE 6-4

AID SPLIT DIAGRAM
JUDGE 2

1. $Y = A_1 U + E^{(1)}$

$A_1 = .48 \quad N_1 = 224$

2. $Y = A_3 X_{16}^{(3,4)} + A_2 X_{16}^{(1,2)} + E^{(2)}$

$A_3 = .85 \quad N_3 = 108$

$A_2 = .13 \quad N_2 = 116$

3. $Y = A_5 X_{10}^{(2,3,4)} + A_4 X_{10}^{(1)} + A_2 X_{16}^{(1,2)} + E^{(3)}$

$A_5 = .91 \quad N_5 = 91$

$A_4 = .53 \quad N_4 = 17$

4. $Y = A_5 X_{10}^{(2,3,4)} + A_4 X_{10}^{(1)} + A_7 X_{16}^{(2)} + A_6 X_{16}^{(1)} + E^{(4)}$

$A_7 = .38 \quad N_7 = 40$

$A_6 = .00 \quad N_6 = 76$

5. $Y = A_5 X_{10}^{(2,3,4)} + A_4 X_{10}^{(1)} + A_9 X_{10}^{(4,3)} + A_8 X_{10}^{(2,1)} + A_6 X_{16}^{(1)} + E^{(5)}$

$A_9 = .68 \quad N_9 = 19$

$A_8 = .10 \quad N_8 = 21$

6. $Y = \underbrace{A_{11} X_{11}^{(3)}}_{SS-5} + \underbrace{A_{10} X_{11}^{(2,1)}}_{SS-6} + \underbrace{A_4 X_{10}^{(1)}}_{SS-4} + \underbrace{A_9 X_{10}^{(4,3)}}_{SS-3} + \underbrace{A_8 X_{10}^{(2,1)}}_{SS-2} + \underbrace{A_6 X_{16}^{(1)}}_{SS-1} + E^{(6)}$

$A_{11} = .97 \quad N_{11} = 68$

$A_{10} = .74 \quad N_{10} = 23$

FIGURE 6-5

AID SPLIT DIAGRAM
JUDGE 3

1. $Y = A_1 U + E^{(1)}$

$A_1 = .46 \quad N_1 = 224$

2. $Y = A_3 X_{16}^{(3,4)} + A_2 X_{16}^{(1,2)} + E^{(2)}$

$A_3 = .86 \quad N_3 = 108$

$A_2 = .09 \quad N_2 = 116$

3. $Y = A_5 X_{11}^{(2,3)} + A_4 X_{11}^{(1)} + A_2 X_{16}^{(1,2)} + E^{(3)}$

$A_5 = .92 \quad N_5 = 93$

$A_4 = .47 \quad N_4 = 15$

4. $Y = A_5 X_{11}^{(2,3)} + A_4 X_{11}^{(1)} + A_7 X_{16}^{(2)} + A_6 X_{16}^{(1)} + E^{(4)}$

$A_7 = .25 \quad N_7 = 40$

$A_6 = .01 \quad N_6 = 76$

5. $Y = \underbrace{A_5 X_{11}^{(2,3)}}_{\text{SS-5}} + \underbrace{A_4 X_{11}^{(1)}}_{\text{SS-4}} + \underbrace{A_9 X_{10}^{(3,2,4)}}_{\text{SS-3}} + \underbrace{A_8 X_{10}^{(1)}}_{\text{SS-2}} + \underbrace{A_6 X_{16}^{(1)}}_{\text{SS-1}} + E^{(5)}$

$A_9 = .45 \quad N_9 = 20$

$A_8 = .05 \quad N_8 = 20$

FIGURE 6-6

AID SPLIT DIAGRAM
JUDGE 4

1. $Y = A_1 U + E^{(1)}$

$A_1 = .53 \quad N_1 = 224$

2. $Y = A_3 X_{16}^{(2,3,4)} + A_2 X_{16}^{(1)} + E^{(2)}$

$A_3 = .78 \quad N_3 = 148$

$A_2 = .04 \quad N_4 = 76$

3. $Y = A_5 X_{16}^{(3,4)} + A_4 X_{16}^{(2)} + A_2 X_{16}^{(1)} + E^{(3)}$

$A_5 = .87 \quad N_5 = 108$

$A_4 = .55 \quad N_4 = 40$

4. $Y = A_7 X_{11}^{(2,3)} + A_6 X_{11}^{(1)} + A_4 X_{16}^{(2)} + A_2 X_{16}^{(1)} + E^{(4)}$

$A_7 = .92 \quad N_7 = 93$

$A_6 = 53 \quad N_6 = 15$

5. $Y = \underbrace{A_7 X_{11}^{(2,3)}}_{\text{SS-5}} + \underbrace{A_6 X_{11}^{(1)}}_{\text{SS-4}} + \underbrace{A_8 X_{13}^{(1,2)}}_{\text{SS-3}} + \underbrace{A_9 X_{13}^{(3)}}_{\text{SS-2}} + \underbrace{A_2 X_{16}^{(1)}}_{\text{SS-1}} + E^{(5)}$

$A_8 = .27 \quad N_8 = 70$

$A_9 = .15 \quad N_9 = 13$

FIGURE 6-7

AID SPLIT DIAGRAM
JUDGE 5

1. $Y = A_1 U + E^{(1)}$

$A_1 = .50 \quad N_1 = 224$

2. $Y = A_3 X_{16}^{(3,4)} + A_2 X_{16}^{(1,2)} + E^{(2)}$

$A_3 = .81 \quad N_3 = 108$

$A_2 = .21 \quad N_2 = 116$

3. $Y = A_3 X_{16}^{(3,4)} + A_5 X_{16}^{(2)} + A_4 X_{16}^{(1)} + E^{(3)}$

$A_5 = .45 \quad N_5 = 40$

$A_4 = .08 \quad N_4 = 76$

4. $Y = A_7 X_{11}^{(3)} + A_6 X_{11}^{(1,2)} + A_5 X_{16}^{(2)} + A_4 X_{16}^{(1)} + E^{(4)}$

$A_7 = .95 \quad N_7 = 74$

$A_6 = .50 \quad N_6 = 34$

5. $Y = \underbrace{A_{19} X_5^{(2,3)}}_{\text{SS-5}} + \underbrace{A_{18} X_5^{(1)}}_{\text{SS-4}} + \underbrace{A_6 X_{11}^{(1,2)}}_{\text{SS-3}} + \underbrace{A_5 X_{16}^{(2)}}_{\text{SS-2}} + \underbrace{A_4 X_{16}^{(1)}}_{\text{SS-1}} + E^{(5)}$

$A_{19} = 1.0 \quad N_{19} = 54$

$A_{18} = .80 \quad N_{18} = 20$

FIGURE 6-8

within the subspaces, first-order models tended to cross-validate better. The coefficients for the local models were obtained by the iterative procedure given in Chapter IV. The local models for the subspaces shown in Figures 6-9 through 6-13, are given in Tables 6-7 through 6-11.

SUBSPACES: JUDGE 1

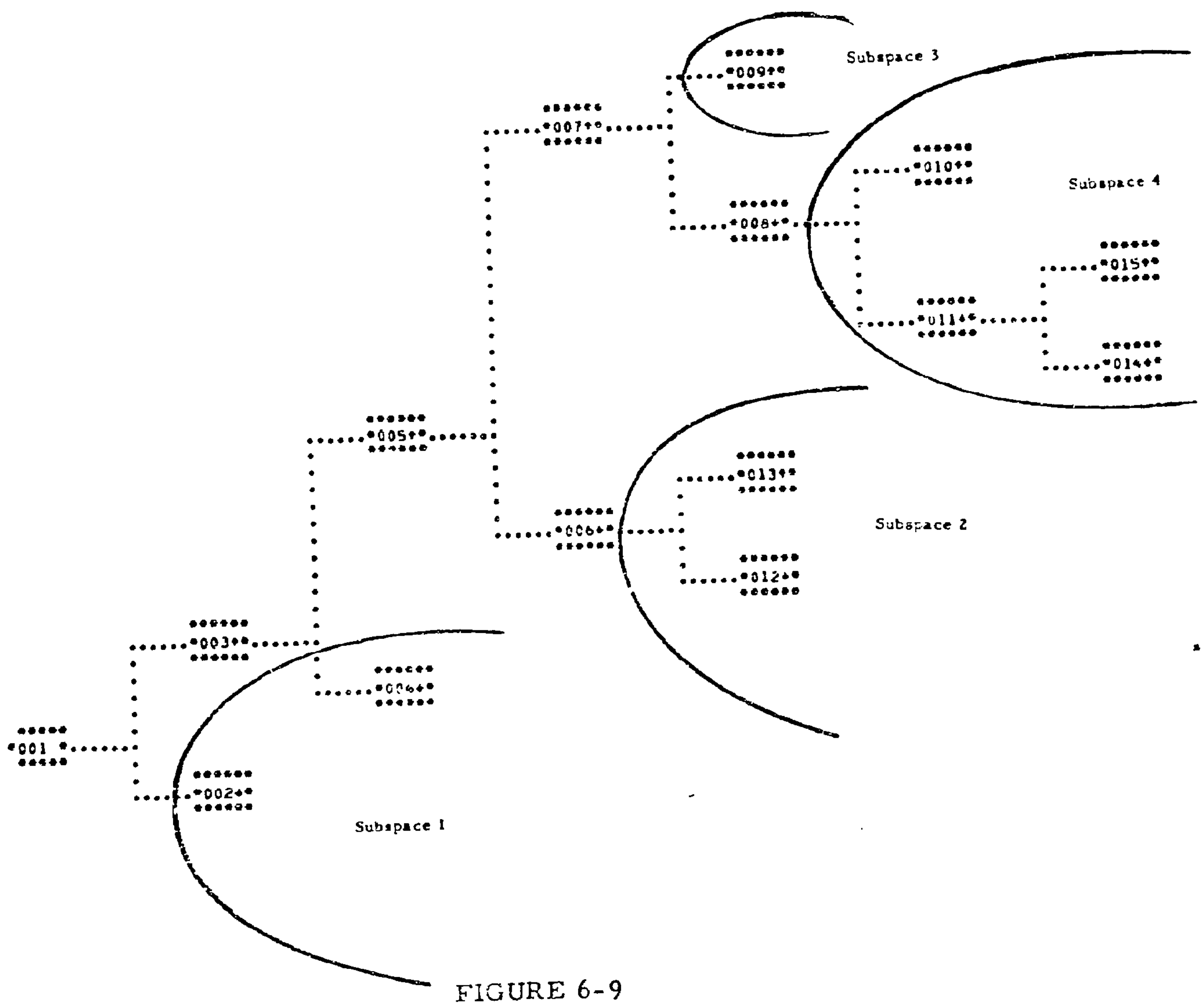


FIGURE 6-9

Table 6-7

Local Model Coefficients: Judge 1

Constant = -.6618

Subspace 1: (GM1=1)	"Poor" Credit Rating or "Other" Financial Ref.
local model 1:	PS = .6687 (GM1) + constant
Subspace 2: (GM2=1)	"No Bank" and not in subspace 1 (SS1)
local model 2:	PS = .3701 (Credit Rating) + .0722 (Phone) + .0600 (ResCat) + constant
Subspace 3: (GM3=1)	"Good" Credit Rating and not in SS1 or SS2
local model 3:	PS = 1.639 (GM3) + constant
Subspace 4: (GM4=1)	Not in SS1, SS2, or SS3
local model 4:	PS = .2258 (Local) + .2044 (ResCat) + .0901 (Type Employment) + .0818(CP1) + .3312 (GM4) + constant
where:	CP1 $\begin{cases} 1 \text{ if Restime} < 2 \text{ and ResCat} < 3 \\ 0 \text{ otherwise} \end{cases}$
Best Cut Score:	.58-.61
Hit Rate:	.929

SUBSPACES: JUDGE 2

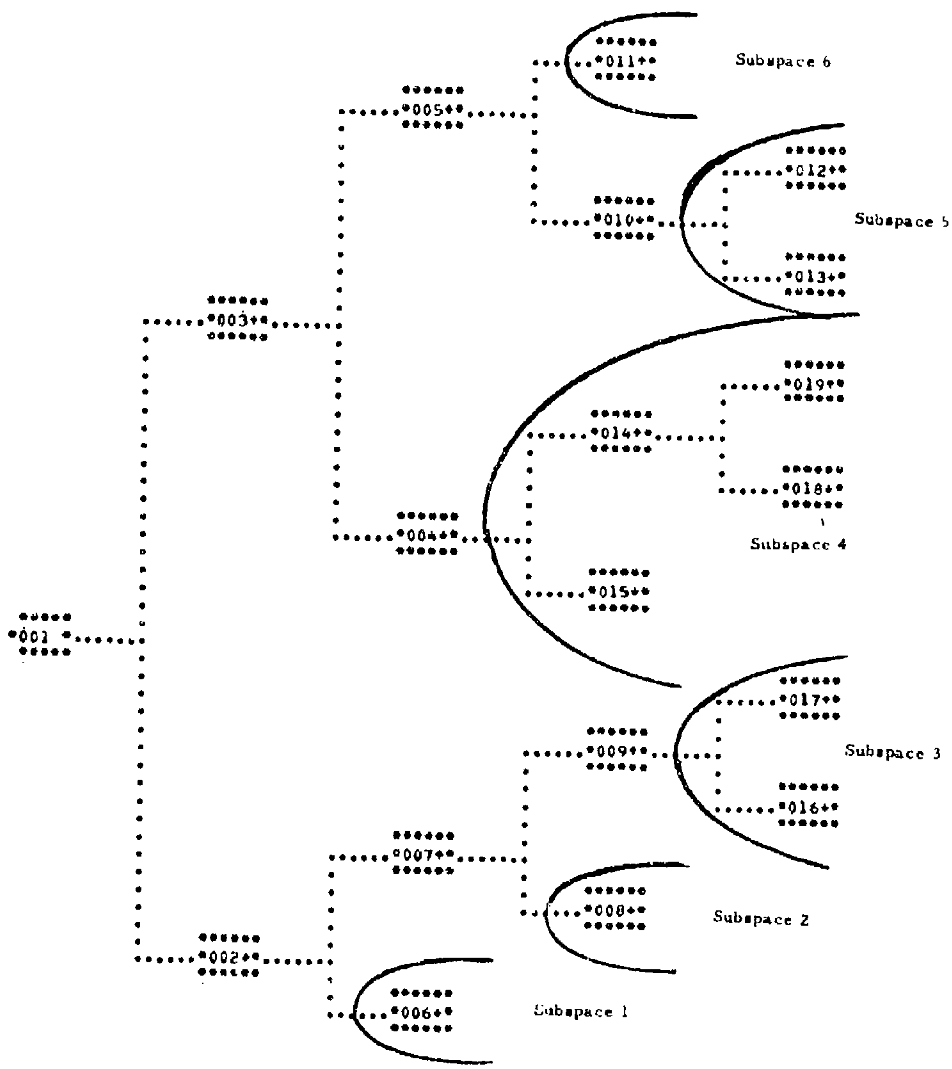


FIGURE 6-10

Table 6-8

Local Model Coefficients: Judge 2

Constant = .760

Subspace 1: (GM1=1)	"Poor" Credit Rating
local model 1:	PS = -.7619 (GM1) + constant
Subspace 2: (GM2=1)	"Unrated" Credit Rating and No Checking Account
local model 2:	PS = -.6642 (GM2) + constant
Subspace 3: (GM3=1)	"Unrated" Credit Rating with at least "Checking" Account
local model 3:	PS = .0910 (Phone) -.3271 (Jobtime) + .0879 (Loan Term) + .1632 (Financial Ref.) -.0341 (GM3) + constant
Subspace 4: (GM4=1)	Credit Rating "Medium" or "Good" and "No-Bank" Account
local model 4:	PS = -.1639 (Type Employment) -.1278 (Income) + .1602 (Financial Ref.) + constant
Subspace 5: (GM5=1)	Credit Rating of "Good" or "Medium", "Savings" or "Checking" and "Major" Financial Reference
local model 5:	PS = .211 (GM5) + constant
Subspace 6: (GM6=1)	Not in any of the other subspaces
local model 6:	PS = .2471 (ResCat) + .0159 (Loan Term) -.3475 Loan Amt. -.0099 (GM6) + constant
Best Cut Score:	.46- 49
Hit Rate:	.929

SUBSPACES: JUDGE 3

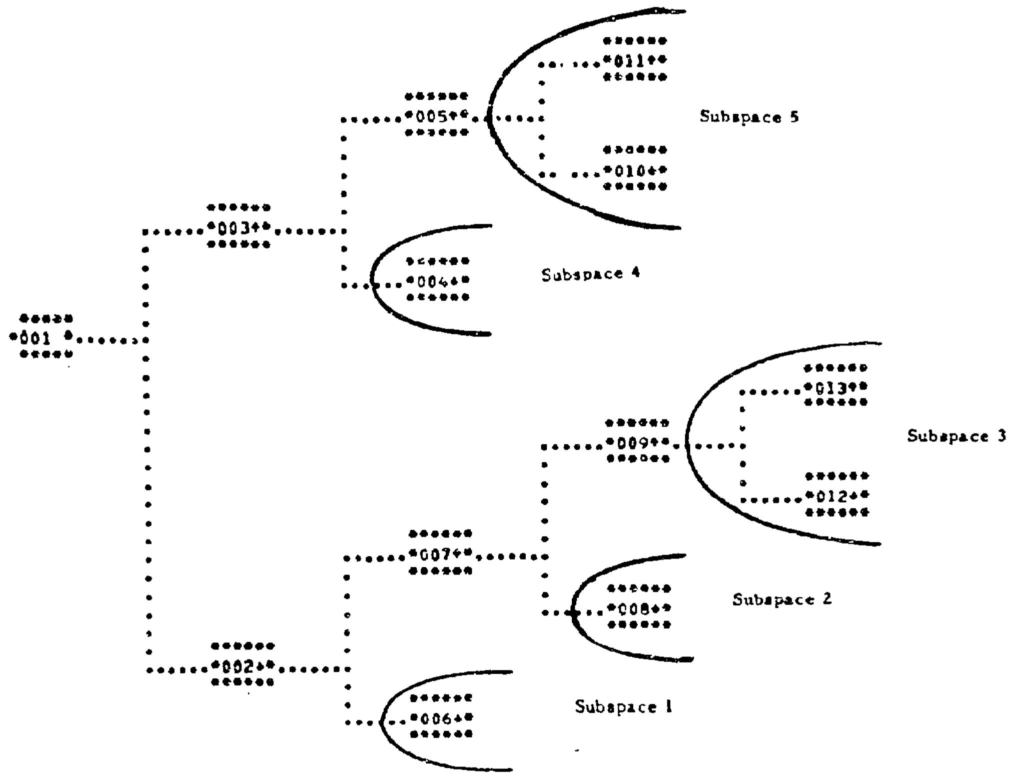


FIGURE 6-11

Table 6-9

Local Model Coefficients: Judge 3

Constant = -.2012

Subspace 1:	"Poor" Credit Rating
(GM1=1)	
local model 1:	PS = .2106 (GM1) + constant
Subspace 2:	"Unrated" Credit Rating and "No Bank"
(GM2=1)	
local model 2:	PS = .2426 (GM2) + constant
Subspace 3:	"Unrated" Credit Rating and "Checking" or
(GM3=1)	"Savings"
local model 3:	PS = .0079 (Phone) + .1099 (Fin. Ref.) + .2973 (Necessity) + constant
Subspace 4:	"Medium" or "Good" Credit Rating and "Other"
(GM4=1)	Financial References
local model 4:	PS = .1165 (Bank Acct.) + .0489 (Credit Rating) + .3720 (CP1) + constant
where:	CP1 $\begin{cases} = 1 & \text{if Jobtime is} > 2 \text{ years} \\ = 0 & \text{otherwise} \end{cases}$
Subspace 5:	"Medium" or "Good" Credit Rating and "Minor"
(GM5=1)	or "Major" Financial References
local model 5:	PS = .2323 (Phone) + .0571 (Type Employment) + .1641 (Fin. Ref.) + .0180 (Credit Rating) + constant
Best Cut Score:	.53-.65
Hit Rate:	.92

SUBSPACES: JUDGE 4

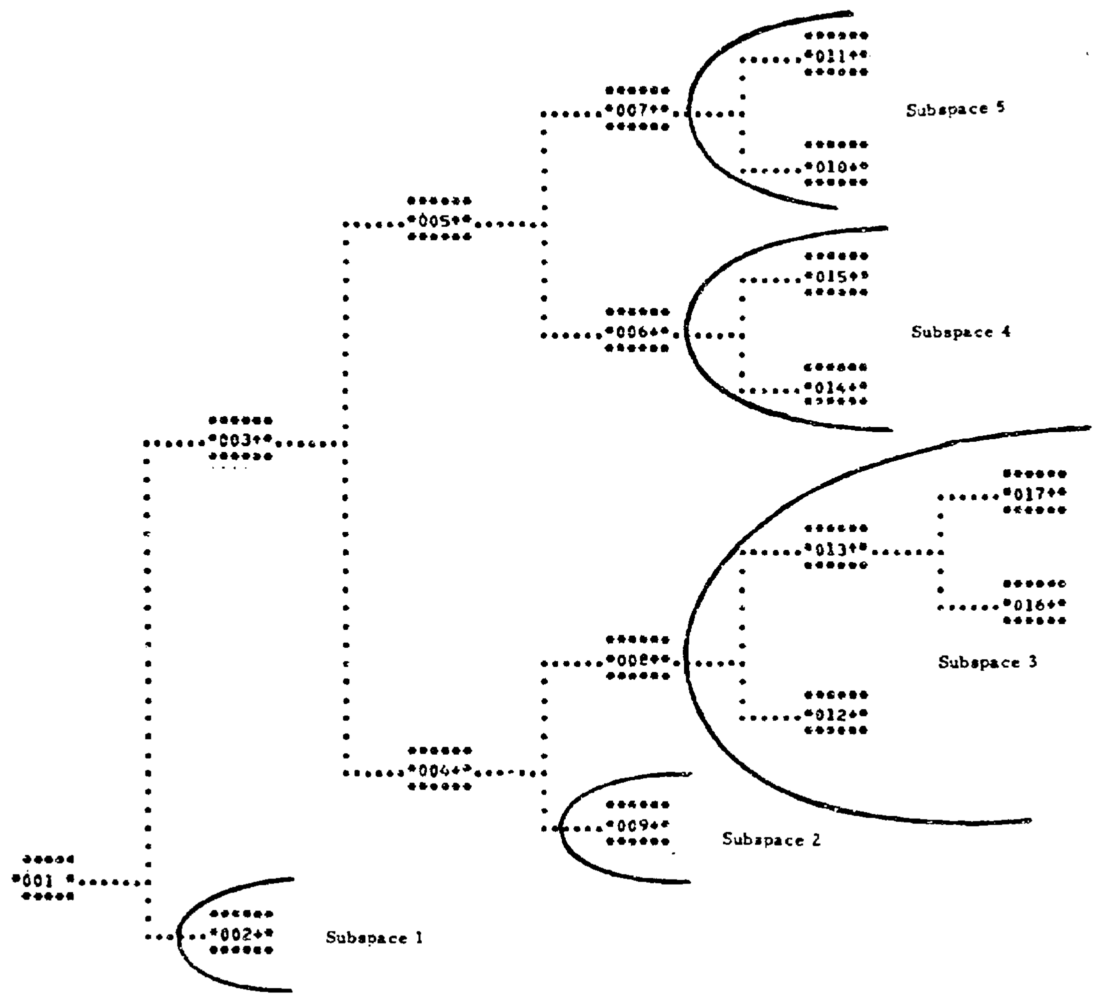


FIGURE 6-12

Table 6-10

Local Model Coefficients: Judge 4

Constant = .0546

Subspace 1:	"Poor" Credit Rating
(GM1=1)	
local model 1:	PS = -.0199 (GM1) + constant
Subspace 2:	"Unrated" Credit Rating and "3 year" Loan
(GM2=1)	Term
local model 2:	PS = .0600 (GM2) + constant
Subspace 3:	"Unrated" Credit Rating and "2 years" or less
(GM3=1)	Loan Term
local model 3:	PS = -.1647 (Local) + .1537 (Income) + .0374 (Bank Acct.) + .2132 (Fin. Ref.) + constant
Subspace 4:	"Medium" or "Good" Credit Rating and "Other"
(GM4=1)	Fin. Ref.
local model 4:	PS = .1479 (Type Employment) + .0761 (GM4) + constant
Subspace 5:	"Medium" or "Good" Credit Rating and "Minor"
(GM5=1)	or "Major" Financial Reference
local model 5:	PS = .4337 (Phone) + .0553 (ResCat) + .0692 (Credit Rating) -.3508 (GM5) + constant
Best Cut Score:	.53-.57
Hit Rate:	.92

SUBSPACES: JUDGE 5

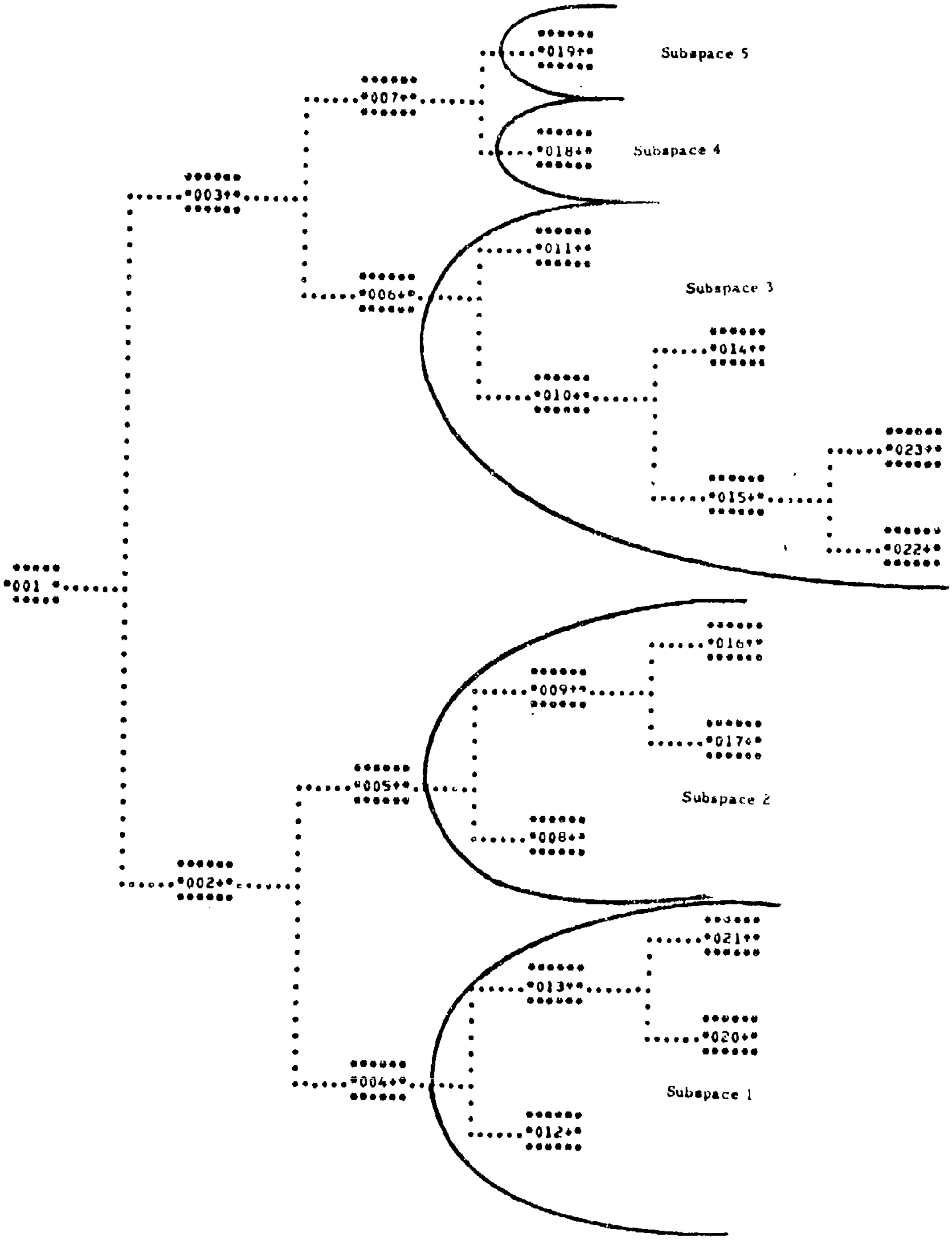


FIGURE 6-13

Table 6-11

Local Model Coefficients: Judge 5

Constant = -.5003

Subspace 1:	"Poor" Credit Rating
(GM1=1)	
local model 1:	PS = .5132 (GM1) + .6865 (CP1) + constant
where:	CP1 $\begin{cases} = 1 & \text{Jobtime} = 5 \text{ years and "Major" Fin.Ref.} \\ = 0 & \text{Otherwise} \end{cases}$
Subspace 2:	"Unrated" Credit Rating
(GM2=1)	
local model 2:	PS = -.2696 (GM2) + .2144 (Marital)+.0452 (Jobtime) + .0927 (Fin. Ref.) -.0707 (Phone) + .2936 (CP2)(Phone) + constant
where:	CP2 $\begin{cases} = 1 & \text{if "checking" or savings for Bank Acct.} \\ = 0 & \end{cases}$
Subspace 3:	"Medium" or "Good" Credit Rating and "Minor" or "Other" Financial Ref.
(GM3=1)	
local model 3:	PS = .1153 (Local) -.033(Phone) +.1616(Restime) +.0950 (Credit Rating)(Fin. Ref.) + .0220 (ResCat) + constant
Subspace 4:	"Medium" or "Good" Credit Rating, "Major" Fin. Ref. and "Less than 2 years" Jobtime
(GM4=1)	
local model 4:	PS = .5221(Phone) + .1260 (Type Employment) -.0131(GM4) + constant
Subspace 5:	"Medium" or "Good" Credit Rating, "Major" Fin. Ref. and "Over 2 years" Jobtime
(GM5=1)	
local model 5:	PS = 1.5011 (GM5) + constant
Best Cut Score:	.91
Hit Rate:	.44-.47

Section IV:

Comparison of Models:

For the first-order, second-order, and local models, the model coefficients shown in Tables 6-5 through 6-11 were used to produce a prediction hit/miss table of the type shown in Exhibits 16 through 18, Appendix B. This provided the data to construct error curves reflecting the number of false approvals versus false denials for each possible cut score between .0 and 1.0. The curves in Figures 6-14 through 6-18 reflect the profiles of these error rates for each model type and each judge. The points plotted on the curves reflect the error rates at cut scores where both the level of false approvals and the level of false denial change. Thus, each point on the curve is a local optimum cut score since it reflects the minimum false approval rate attainable at a particular false denial rate, or vice versa. The important thing to note from these curves is that the curves for the local models dominate the curves for the first-order and second-order models; that is, at any given false approval rate, the local model for the judge will result in the same <u>or fewer</u> false denials than his first-order or second-order model.

In this sense the "goodness" of the model is being measured relative to the model's predictive capability over the whole spectrum of possible cut scores and not solely on the model's maximum hit rate.

Comparison of Judges:

Comparison of the judges' policies solely on the basis of

ERROR CURVES FOR JUDGE I

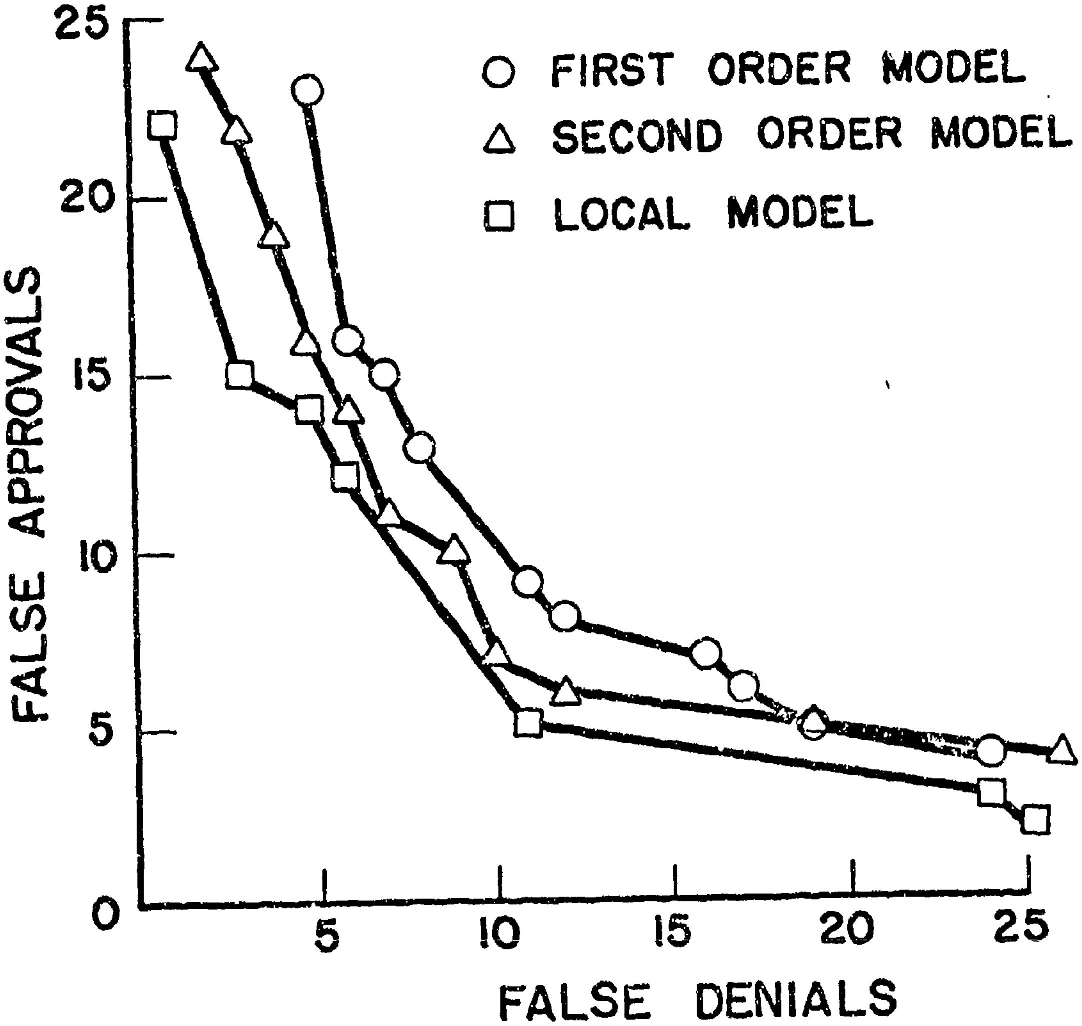


FIGURE 6-14

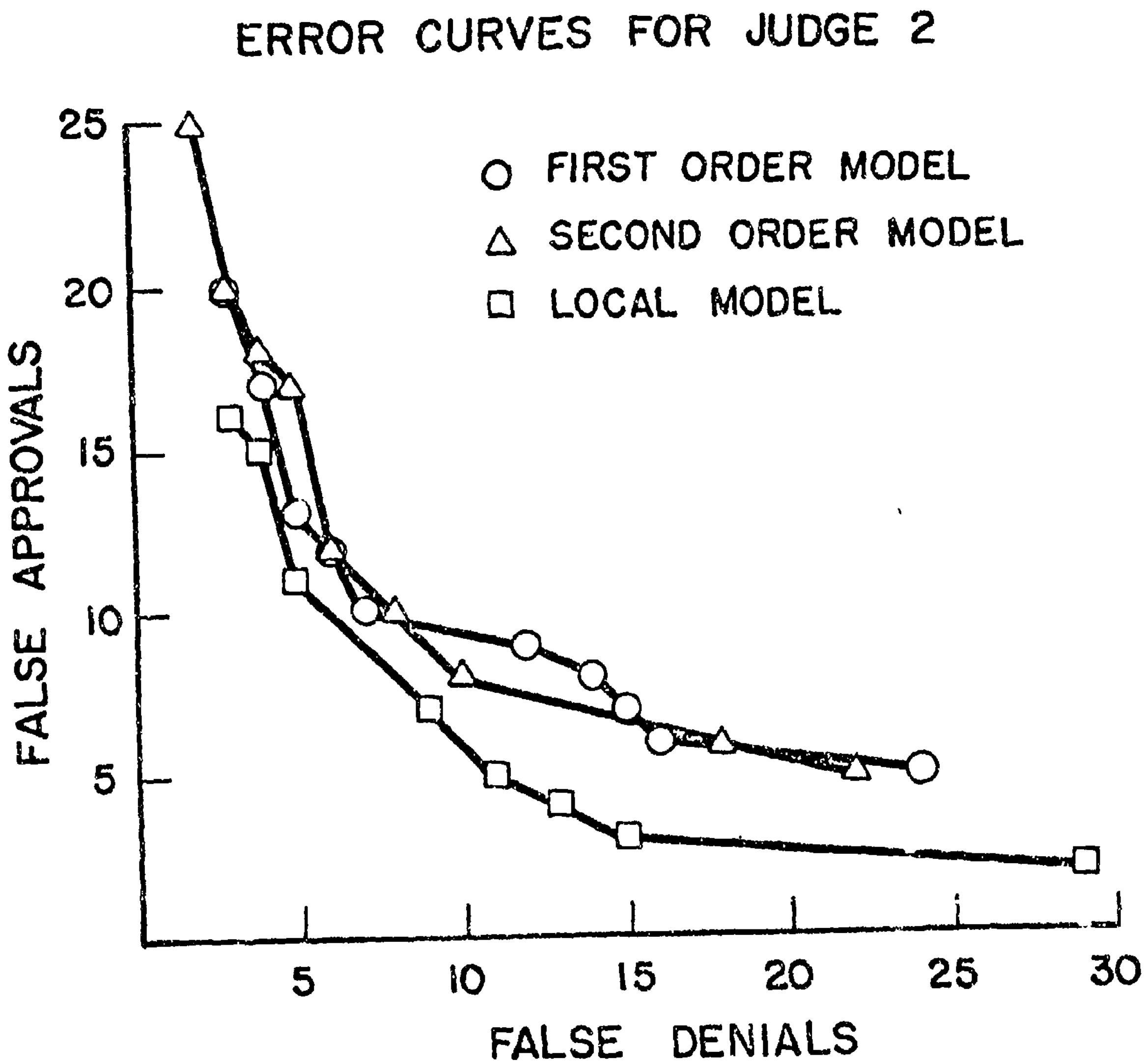

ERROR CURVES FOR JUDGE 2
25
20
15
10
5
FALSE APPROVALS
FIRST ORDER MODEL
SECOND ORDER MODEL
LOCAL MODEL
5
10
15
20
25
30
FALSE DENIALS

FIGURE 6-15

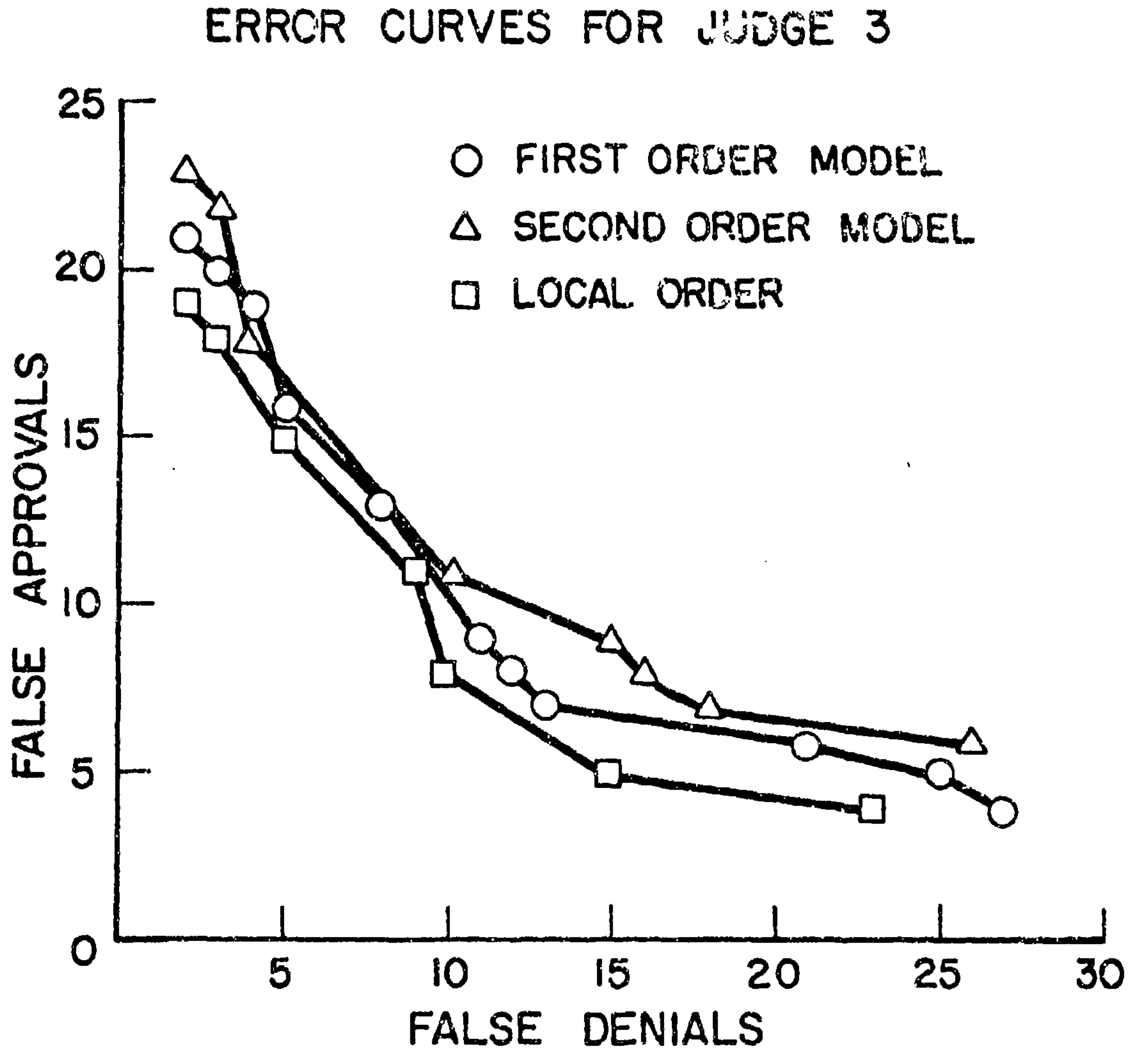

ERROR CURVES FOR JUDGE 3
FIRST ORDER MODEL
SECOND ORDER MODEL
LOCAL ORDER
FALSE APPROVALS
FALSE DENIALS
0
5
10
15
20
25
30

FIGURE 6-16

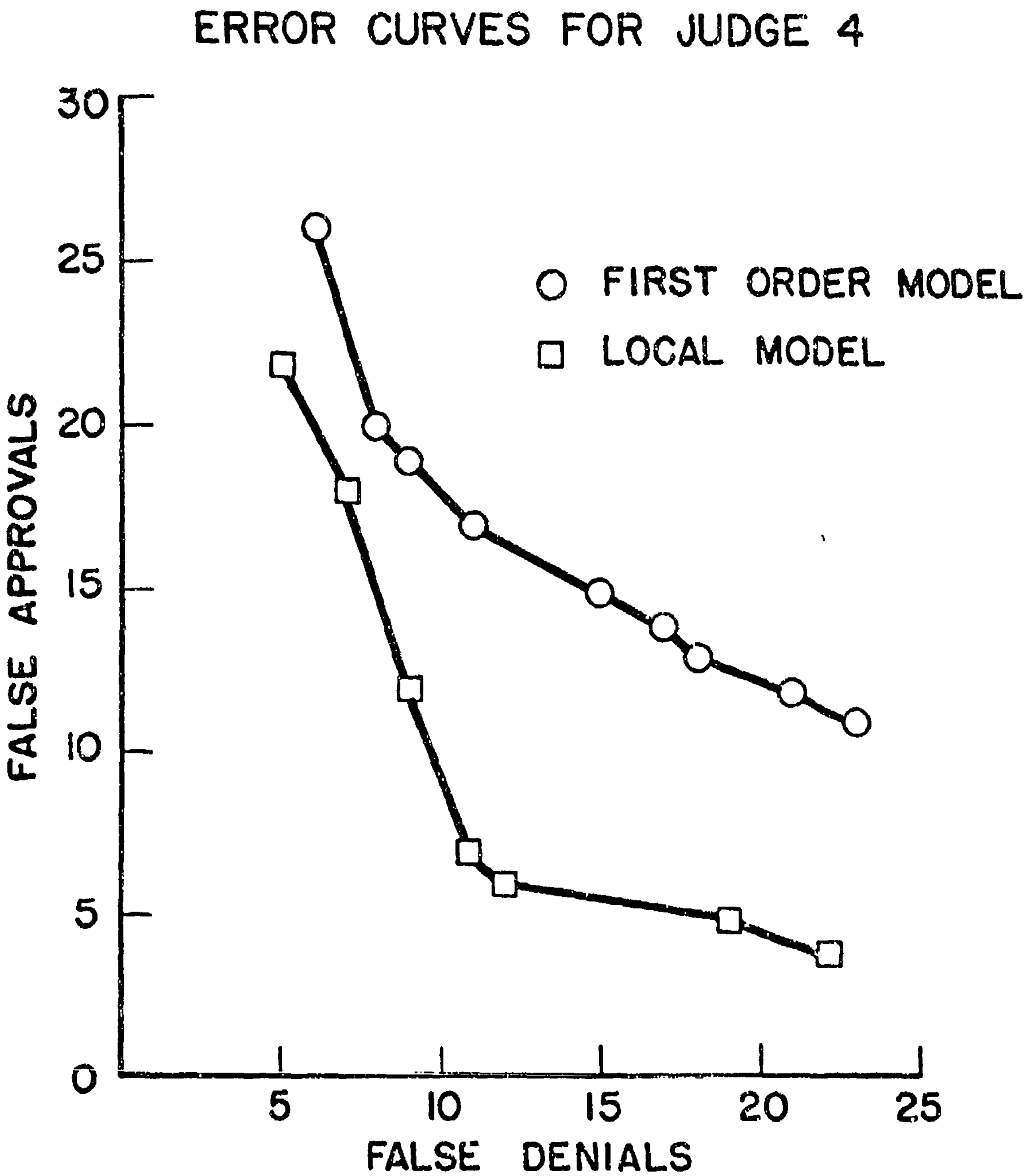

ERROR CURVES FOR JUDGE 4
FIRST ORDER MODEL
LOCAL MODEL
FALSE APPROVALS
FALSE DENIALS
0
5
10
15
20
25
30
5
10
15
20
25

FIGURE 6-17

ERROR CURVES FOR JUDGE 5

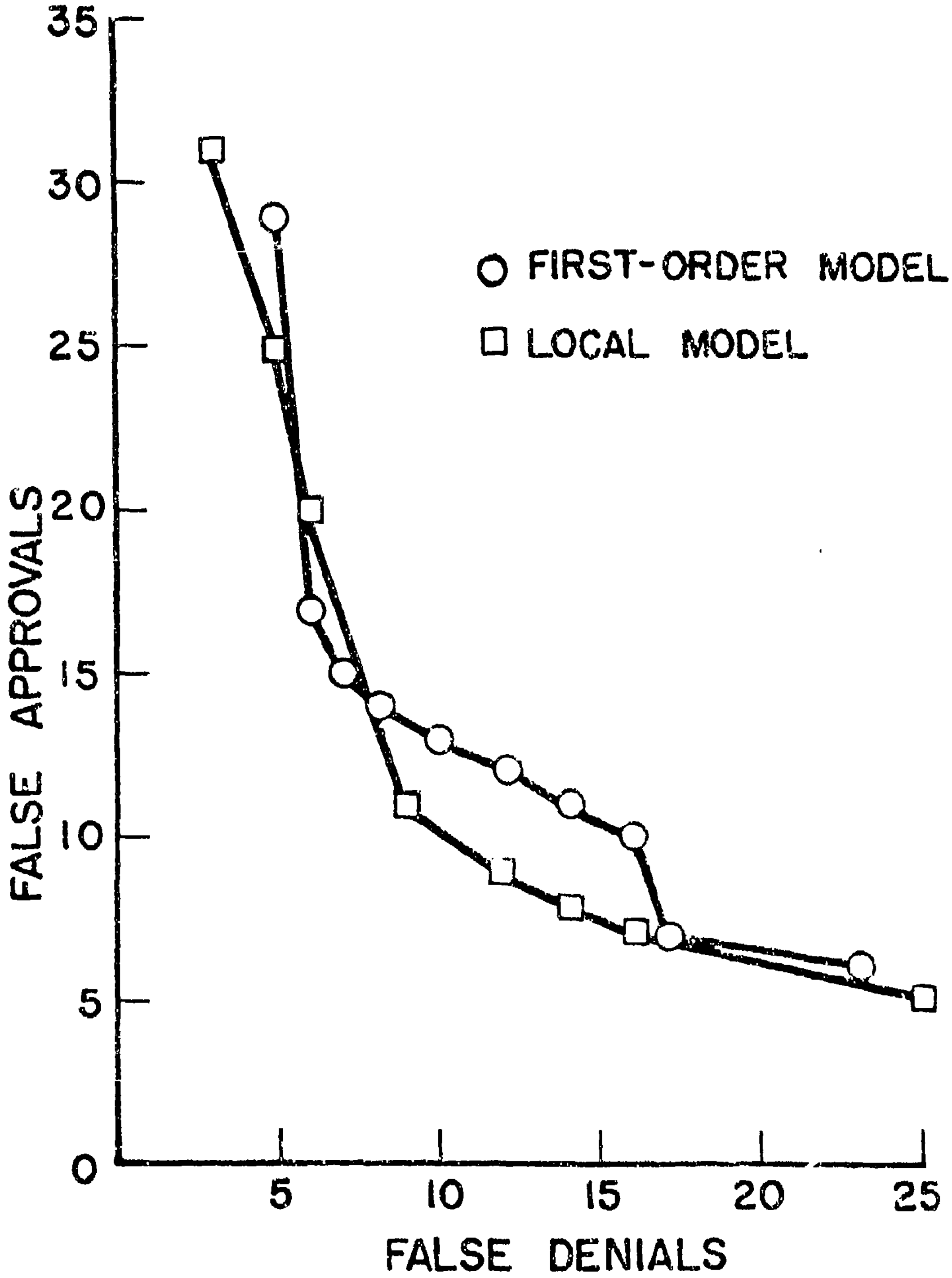


FIGURE 6-18

the magnitude of the coefficients for the predictors is not very informative when the predictors are correlated, as is the case in this decision process. Models with greatly different coefficients can result in very nearly the same set of decisions. For example, Table 6-12 presents the similarity of the decisions of the five judges in terms of the percentage of matches. This data indicates that all of the judges' decisions are the same between 87 and 91 percent, yet Table 6-5 indicates a fairly wide disparity among the coefficients. Note that for Judge 2 and Judge 5, four of the coefficients [illegible]ally have different signs, yet they have the highest match on decisions.

TABLE 6-12

INTERJUDGE SIMILARITY
(percent)

		2	3	4	5	4/5 Vote
1	100	89.3	87	87	87	94
2		100	89.[illegible]	89.8	91	94.5
3			100	88.5	86.4	91.7
4				100	89.6	90.6
5					100	93

Similarly, comparison of the second order (Table 6-6) coefficients does not reveal much useful information.

Comparison of the Aid-Trees for each judge (Figures 6-4 through 6-8) reflects that Judges 1 through 3 are primarily concerned with Credit Rating (X_{16}), Financial References (X_{11}) and Bank Account

(X_{10}). However, Bank Account is not a primary variable for either Judge 4 or Judge 5 while Loan Term (X_{13}) is important to Judge 4 and Time on Job (X_5) is important to Judge 5.

Further, the Aid-Tree identifies a Financial Reference of "Other" as a sufficient condition for denial, regardles of Credit Rating, for Judge 1. Yet Judges 3 and 4 approved 47 percent and 53 percent of applications if they had a "Medium" or "Good" Credit Rating. Careful study of tne Aid-Tree can suggest many such differences. However the question must always be asked, "Is the difference due to policy or is it due to the computational scheme?" During the research it was found that a very useful technique was to present each judge his tree and have him verbalize or explain the splits. This led to exposure and discussion of many subtle differences in opinion among the judges.

Section V:

The Voting and Composite Models:

As was the case in the "without credit" models, a voting strategy of 4/5 appeared to best match the actual approvals and denials. The last column of Table 6-12 reflects that the decision of each of the individual judges matches voting policy better than they match the decisions of any of the other individual judges.

An alternate composite model was also obtained by the more traditional process of clustering first-order models of the individual judges with the EDSTAT-J version of the JAN hierarchial clustering routine. The coefficients for the 4/5 vote and clustered composite models are shown in Table 6-13. Figure 6-19 shows the merge diagram from the clustering process. This diagram reflects the policy comparison technique that has been generally used in the past. It reflects that Judge 1 and Judge 5 have the most similar policies, yet their hit/miss ratio shown in Table 6-12 is one of the lowest.

This paradox is due to the nature of the data and the fact that the criteria for closeness are different. In the clustering approach, the sum of squared errors between actual score and predicted score is the criterion. In the hit/miss approach, the fact of a difference is important, but the magnitude is not (all "miss" magnitudes are 1 and "hits" are 0). The whole point of presenting and discussing the merge diagram is that, with binary decisions, the use of continuous criteria such as the squared multiple correlation coefficient (R^2) or the sum of squared errors is probably not desirable and can lead to vastly different results.

TABLE 6-13

COEFFICIENTS FOR 4/5 VOTE AND COMPOSITE MODELS

Predictors:	Coefficients	
	4/5 Vote	Clustered Composite
Constant	-1.0469	-.8615
Age	- .0268	-.0201
Marital	- .0383	-.0344
Local	.0700	.0610
Phone	.1393	.1053
Jobtime	.0322	.0456
Restime	- .0308	-.0279
ResCat	.0830	.0527
Type Employment	.0151	.0174
Income	- .0225	-.0276
Bank Account	.0655	.0584
Financial Reference	.1538	.1279
Loan Amount	.0072	.0128
Loan Term	- .0585	-.0571
Necessity	.0061	-.0146
Equity	.0066	-.0093
Credit Rating	.2647	.2622
Best Cut Score:	.46	.27
Hit Rate:	.94	.92

Since the 4/5 vote policy model actually represented a set of substitute responses for the actual decisions or the individual judge's decisions, it was used to build an Aid-Tree (Exhibit 13, Appendix B) and to define subspaces and local models. Although the superiority of the local model relative to the first-order model still held up in terms of the error curves shown in Figure 6-20, this superiority was not as great as it was in the case of the individual judges. This decrease is understandable since the 4/5 vote policy tends to average the policies and models of the individual judges, and in so doing, obscures the configuralities and idiosyncracies of the individual policies. Thus the power of the local models in modeling the various subspaces is diminished since the boundaries of these subspaces are not as well defined.

MERGE DIAGRAM FROM CLUSTERING PROGRAM

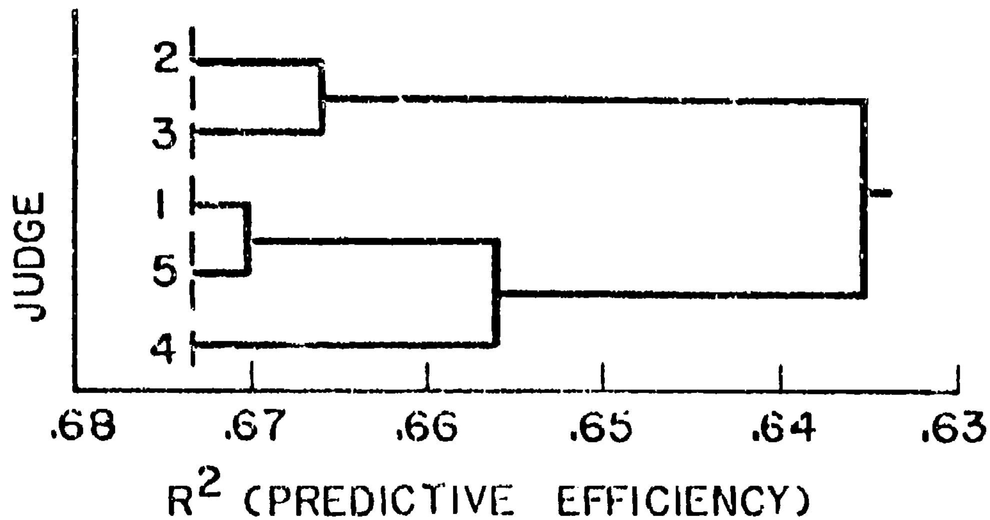


FIGURE 6-19

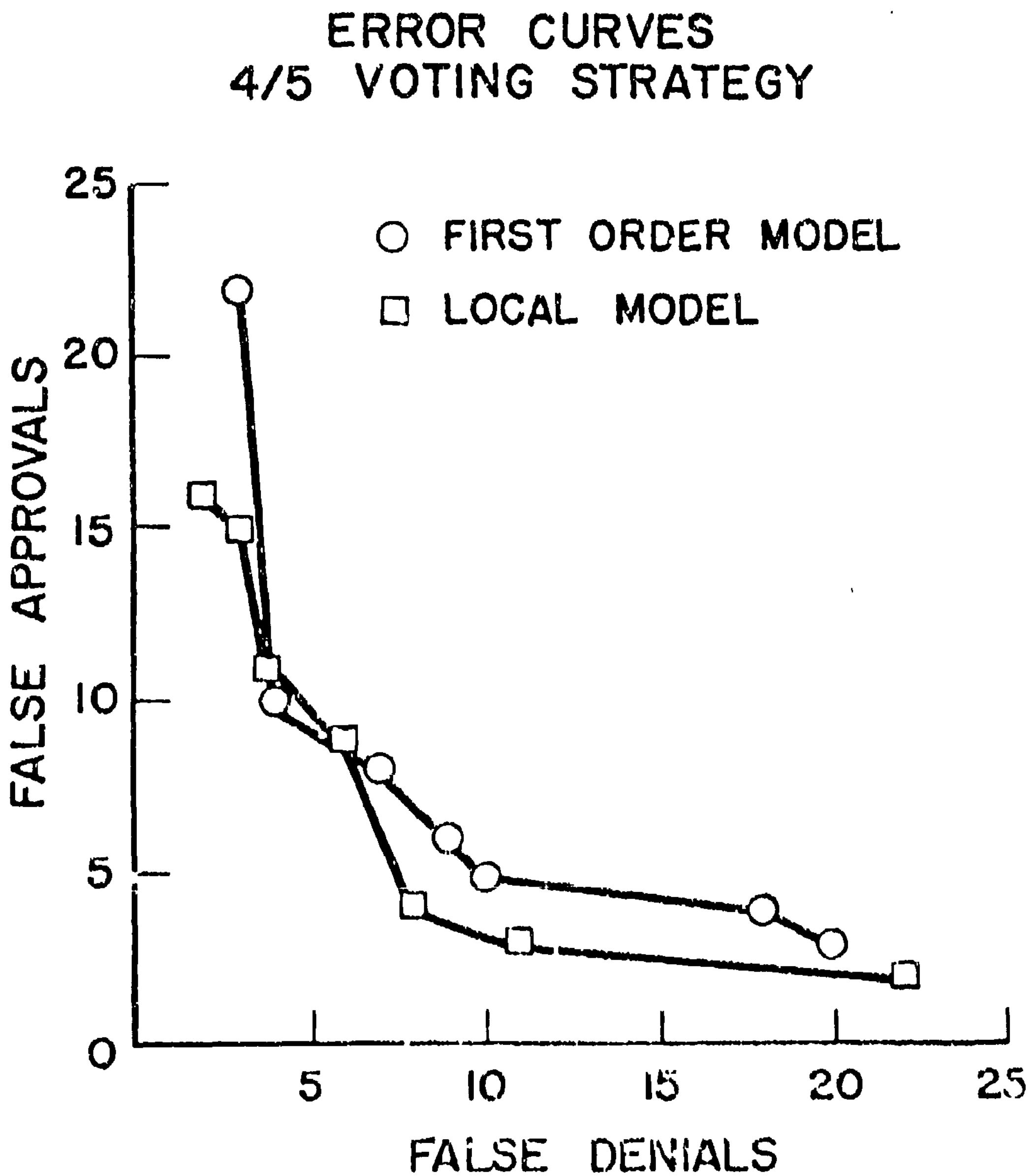

ERROR CURVES
4/5 VOTING STRATEGY
FIRST ORDER MODEL
LOCAL MODEL
FALSE APPROVALS
FALSE DENIALS
0
5
10
15
20
25
5
10
15
20
25

FIGURE 6-20

Section VI.

Results of the Cross-Validation Field Test:

A Shift in the Sample Characteristics and the Policy:

Of the 150 new applications collected during the field test, 75 percent were given loans without the stipulation of "full recourse". This reflects an increase of 30 percent in the approval rate between the 224 applications used in the model building phased and the 150 applications of the cross-validation population. This increase is attributable to two basic factors that are detectable in Table 6-14.

TABLE 6-14

APPROVAL RATE vs. CREDIT RATING CATEGORY

Credit Rating	Original Sample			Cross-Validation Sample		
	#	Percent of Sample	Percent Approval	#	Percent of Sample	Percent Approval
Poor	76	34	0	27	18	7.5
Unequal	39	17	40	27	18	66
Medium	46	22	56	16	11	87
Good	62	27	95	80	53	99
Totals	224	100	45	150	100	75

First, there was an apparent increase in the caliber of the applications relative to the primary decision variable, Credit Rating. Second, there was an increase in the approval rate for all categories of Credit Rating.

There are two probable explanations for the increase in the caliber of the applications:

1) sampling differences between the random sample of 224 and the total July 72 population of applications used for cross-validation

2) an actual shift in the credit credentials of the potential borrower during the month of July 72.[4]

The increased approval rate reflects a change in the relationship between the dealer and the lending institution. This change is attributable to the longer time that they had worked together by the time the cross-validation sample was taken. Whereas the original sample reflected their relationship after only 4 to 5 months of the dealer contract, they had been doing business approximately 14 months at the time of the second sample and the trust and understanding between them was greatly improved. During this period, the volume of business from Dealer B had grown by some 400 percent and the loan officers expressed a willingness to accept somewhat riskier applications in order to keep the dealer happy and continue the expansion of business. This change is especially noticeable in the applications with

---

[4] Discussion with Judge 2 revealed there is a fluctuation of the credit quality that is a function of the general economic conditions. When business is slow, a general decrease of credit quality prevails.

"Unrated" and "Medium" Credit Rating.

"With Credit" Policy Models vs. Decisions on the Codified Predictors:

One result of these shifts was to improve the performance of the policy models and lower the cut scores corresponding to the optimum hit rates. As can be seen in Table 6-15, the cross-validation hit rates for the judges' decisions, based on the codified predictor sets, were generally be. er than the hit rates on the original sample. This

Table 6-15

MAXIMUM HIT-RATES

Equation	Original Hit-Rate	Cross Validation Hit-Rate
Judge 1--first order	.91	.96
Judge 1--second order	.92	.96
Judge 1--local model	.93	.93
Judge 2--first order	.92	.96
Judge 2--second order	.92	.97
Judge 2--local model	.93	.93
Judge 3--first order	.91	.95
Judge 3--second order	.90	.95
Judge 3--local model	.92	.95
Judge 4--first order	.88	.95
Judge 4--local model	.92	.94
Judge 5--first order	.90	.96
Judge 5--local model	.91	.93
4/5 Vote--first order	.94	.97
4/5 Vote--local model	.94	.97

table also reflects that the first-order and second-order models tended to do slightly better than the local models in terms of cross-validation. Given the observed shift in policy, concentrated in the Credit Rating categories of "Unrated" and "Medium", this is not surprising. These are the predictor categories that defined the subspaces over which the local models were the most complex and influential. The superiority of the local models (as shown in Figures 6-14 to 6-19) was mainly due to the better fit of the local hyperplanes over these subspaces. This superior fit would naturally be most affected by perturbation of the policy in those subspaces.

"With Credit" Policy Models vs. Actual Decisions:

The various composite models were cross-validated against the actual approvals and denials that the loan officers had made using the original application forms (Exhibit 1, Appendix B). Exhibits 16, 17, and 19 present the appropriate hit/miss tables and Figure 6-21 shows the error curves for the various policy models. The first-order "4/5" vote and composite models each attained 92 percent accuracy. The local "4/5" vote model was 91 percent accurate. Comparison of the first-order voting model's performance in predicting decisions that were made on the codified predictors with its performance in predicting the actual decisions based on the original application form reflects that there is only a 5 percent difference. This is somewhat less than the difference between using codified predictors and actual applications on the original 224 as reflected by the data from Table 6-3. The decrease probably reflects the judges' adaptation to the use of single category descriptors through their participation in the project. In any event, the difference between the results based on the codified predictors

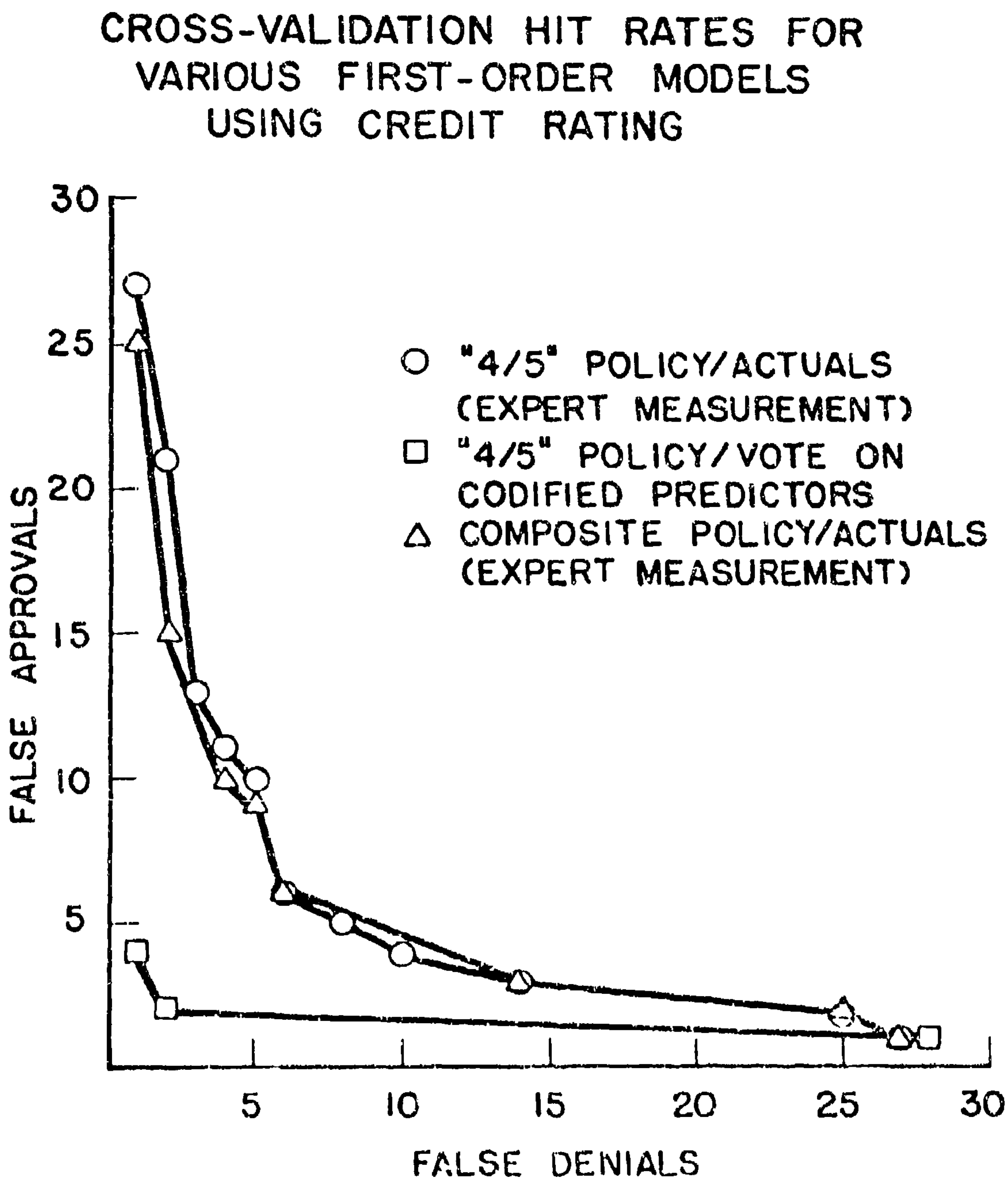

CROSS-VALIDATION HIT RATES FOR
VARIOUS FIRST-ORDER MODELS
USING CREDIT RATING
FALSE APPROVALS
FALSE DENIALS
30
25
20
15
10
5
5
10
15
20
25
30
"4/5" POLICY/ACTUALS (EXPERT MEASUREMENT)
"4/5" POLICY/VOTE ON CODIFIED PREDICTORS
COMPOSITE POLICY/ACTUALS (EXPERT MEASUREMENT)

FIGURE 6-21

and the results based on the natural language applications does not appear to be a significant problem in construction of policy models for this decision process.

Clerical vs. Expert Measurements:

One of the objectives of the field test was to ascertain the capability of the clerical staff to properly codify the application data. To this end, the set of instructions in Exhibit 6, Appendix B, were explained to the clerks and left for their reference in filling out the applications. Their questions were solicited and answered the first day of the field test. After each clerk expressed confidence in her understanding of the instructions, there was no further monitoring of the process of translation from the original application into the codified predictors. However, the judge who made the decision on the application was also asked to fill out the coded application form. This provided a means of comparing the "clerical" measurement with the "expert" measurement. An item-by-item comparison on each of the applications was made and the results are displayed in Table 6-16. The wide disparity in the assessment of the predictors is rather amazing, especially in the case of those predictors which should have required absolutely no judgment. A check of both translations with the original applications revealed the loan officer was generally correct in cases where a difference existed, but not always. Figure 6-22 reflects the effect of this disparity on the decisions predicted by the "4/5" voting model. The lack of a large effect in the predictions is attributable to the fact that the primary predictor in the model is Credit Rating, and this was only available to, and assessed by, the

Table 6-16

EXPERT VS. CLERICAL MEASUREMENT DIFFERENCES ON APPLICATION VARIABLES

Predictor	Cases Assessed Differently*
Age	10
Marital Status	7
Local Family	25
Telephone	12
Job Time	29
Residence Time	29
Residence Category	26
Type Employment	61
Income	52
Bank Account	40
Financial References	69
Loan Amount	32
Loan Term	16
Necessity	18
Equity	35
Credit Rating	Only assessed by loan officers

*At least one item differed on 141 of the 150 applications.

loan officer. In the case of the "tentative approval" model. Credit Rating was not used. A comparison of Figures 6-23 and 6-24 reflects that misassessing the predictors does have a substantial effect. In either case, it would appear that some formal training of the clerical staff would be required to insure that their assessment of the predictors is congruent with that of the loan officers.

"Without Credit" Policy Models vs. Actual Decisions:

The shift toward leniency in the policy of the judges had a positive effect on the observed performance of the policy models used

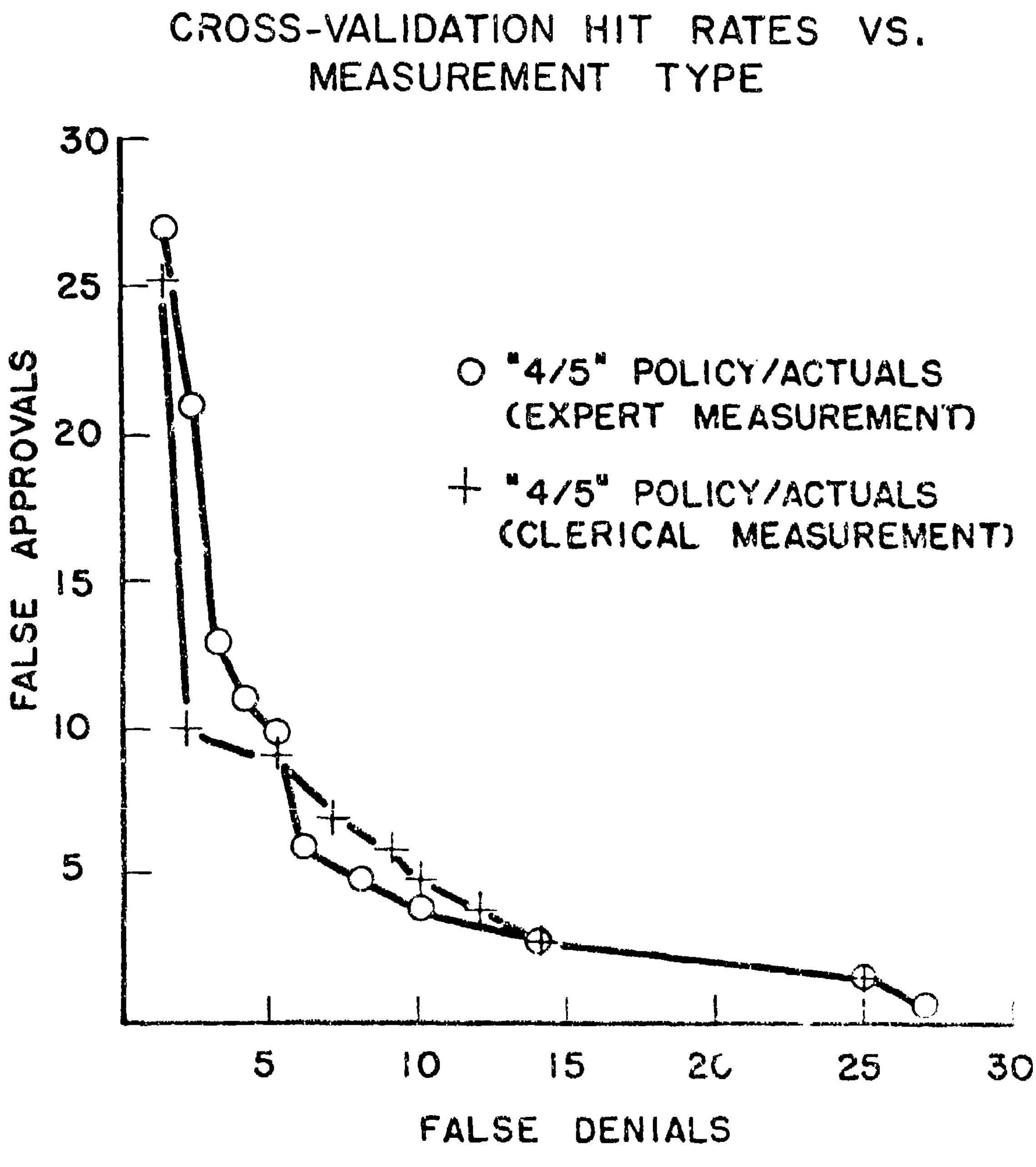

CROSS-VALIDATION HIT RATES VS.
MEASUREMENT TYPE
30
25
20
15
10
5
FALSE APPROVALS
O "4/5" POLICY/ACTUALS
(EXPERT MEASUREMENT)
+ "4/5" POLICY/ACTUALS
(CLERICAL MEASUREMENT)
5
10
15
20
25
30
FALSE DENIALS

FIGURE 6-22

to predict "tentative approval". Figure 6-23 reflects that if everyone scoring .85 or above had been approved, over 48 percent of the applications could have been "tentatively approved" with only a 3 percent error rate. This is well above the 25-30 percent criterion set by the loan officers as a minimum limit for usability. Further, in assessing the mistakes that occurred, it was found that 2 of the 3 false approvals, made by the model at the .85 cut score, had actually been granted under the "full recourse" option. Figure 6-24 indicates the performance of the model with clerically translated predictors is significantly worse and would probably be unacceptable. Figure 6-25 shows that the model which includes "Dings" is somewhat more conservative, but does perform well in terms of error rate. As for implementation, the use of the more conservative model would probably be the better option in order to assure acceptable performance when the overall credit quality is lower.

First-Order vs. Local Models:

The reversal of the first-order and local models in terms of cross-validation hit rate would indicate that the local model is more sensitive to policy changes, especially if the change only affects certain subgroups of the applicants. However, the local model is more robust in the sense that it has a better profile of hit rates over the entire range of application scores, and would be less sensitive to general changes in the policy affecting the entire spectrum of applications. Figure 6-26 presents the profile of hit rates on the cross-validation sample for all possible cut scores. The first-order models enjoy a local superiority at a given cut score; the local models do a better "average" job of predicting the decisions.

RISK CURVES FOR "4/5" POLICY FOR TENTATIVE APPROVAL (EXPERT MEASUREMENT)

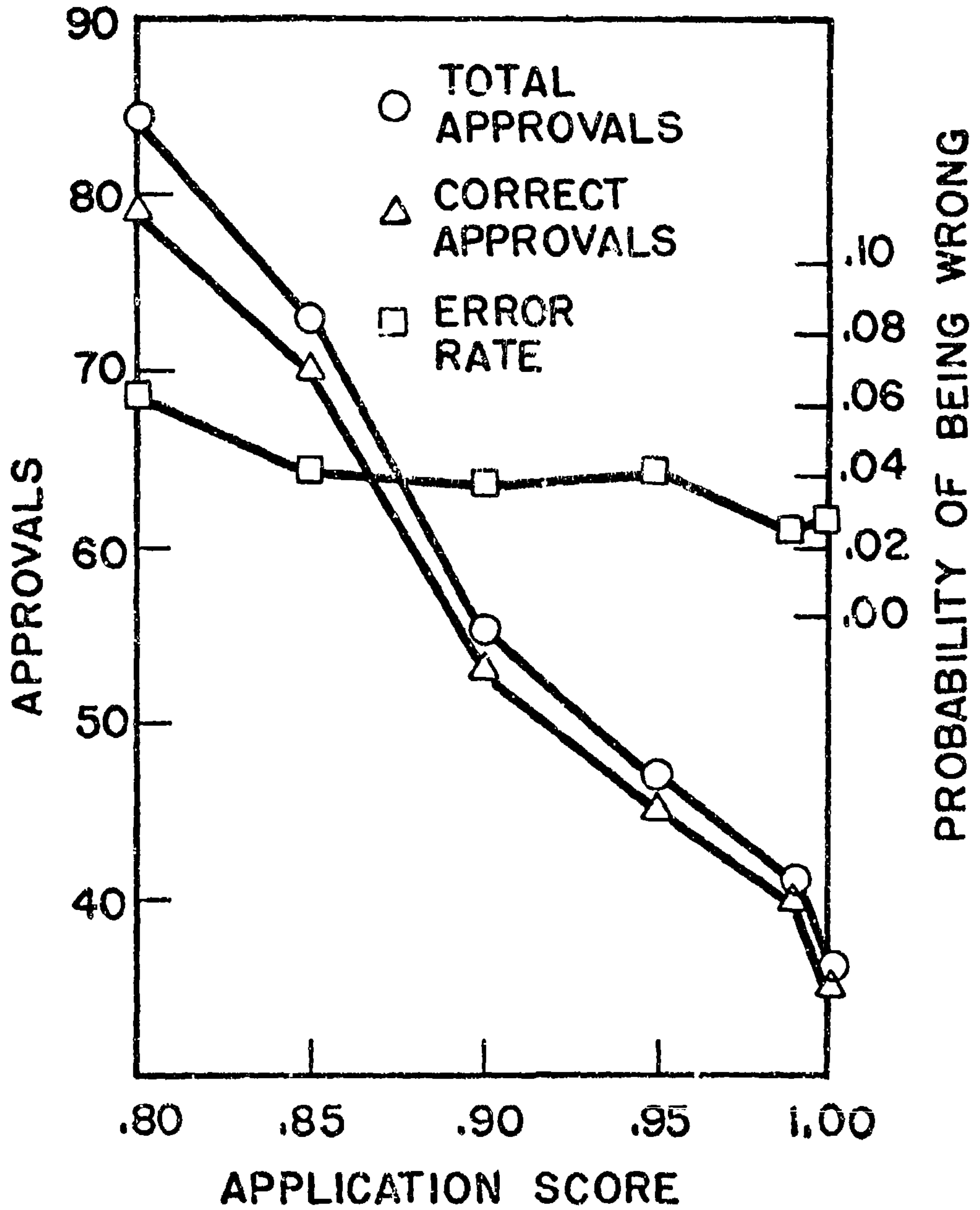


FIGURE 6-23

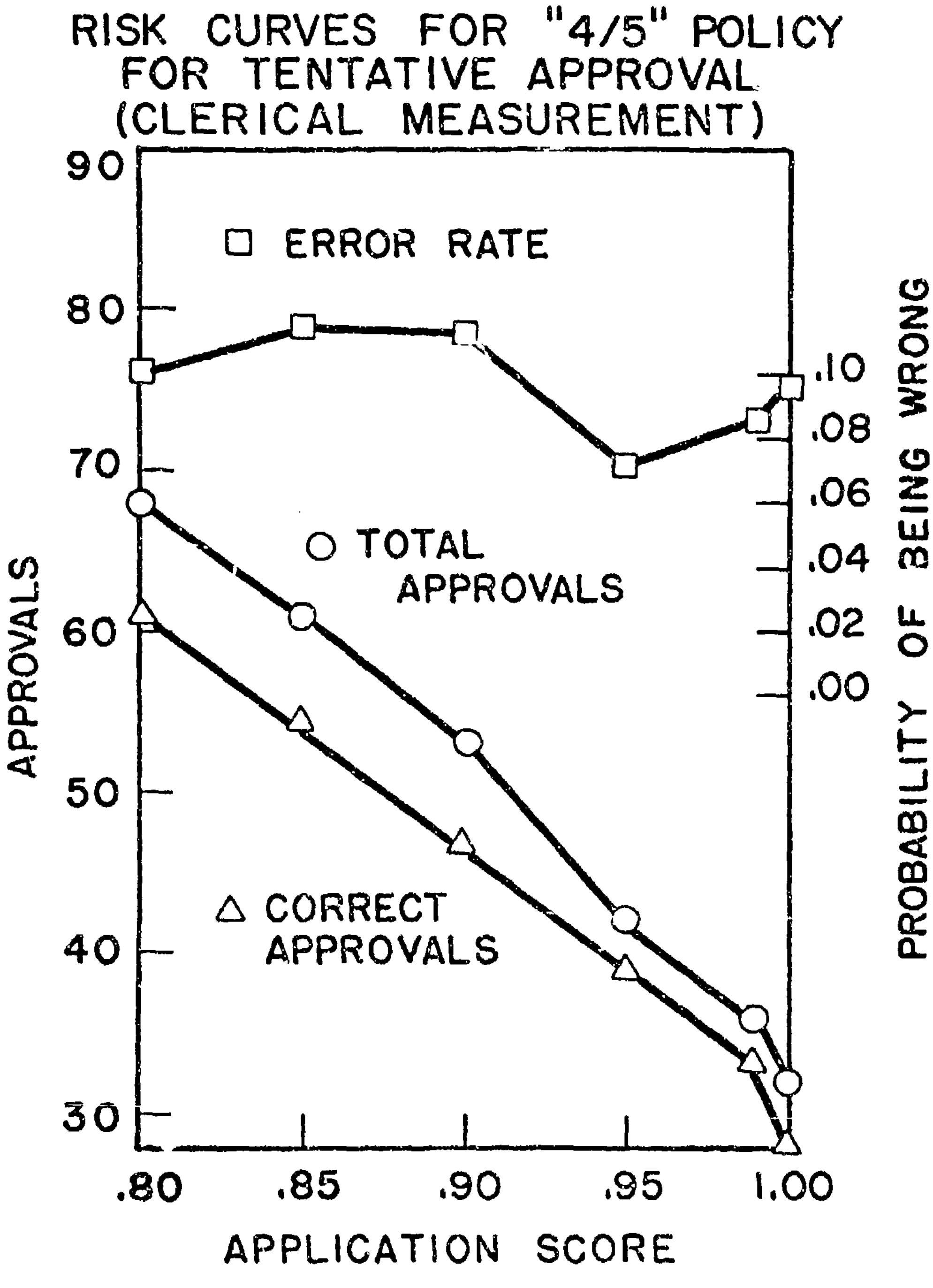

RISK CURVES FOR "4/5" POLICY
FOR TENTATIVE APPROVAL
(CLERICAL MEASUREMENT)
ERROR RATE
TOTAL APPROVALS
CORRECT APPROVALS
APPROVALS
PROBABILITY OF BEING WRONG
APPLICATION SCORE
90
80
70
60
50
40
30
.10
.08
.06
.04
.02
.00
.80
.85
.90
.95
1.00

FIGURE 6-24

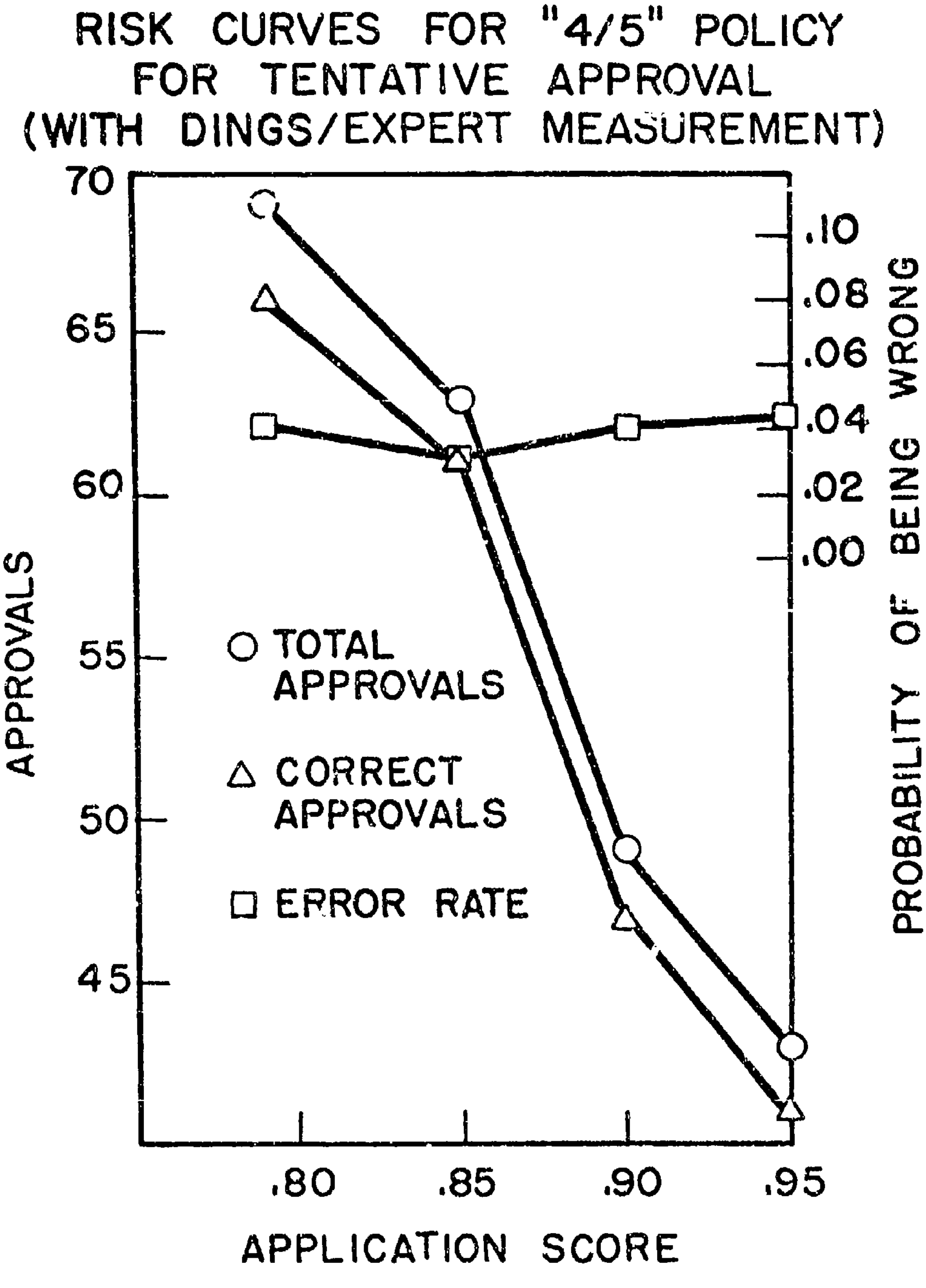

RISK CURVES FOR "4/5" POLICY
FOR TENTATIVE APPROVAL
(WITH DINGS/EXPERT MEASUREMENT)
70
65
60
55
50
45
APPROVALS
.10
.08
.06
.04
.02
.00
PROBABILITY OF BEING WRONG
TOTAL APPROVALS
CORRECT APPROVALS
ERROR RATE
.80
.85
.90
.95
APPLICATION SCORE

FIGURE 6 - 25

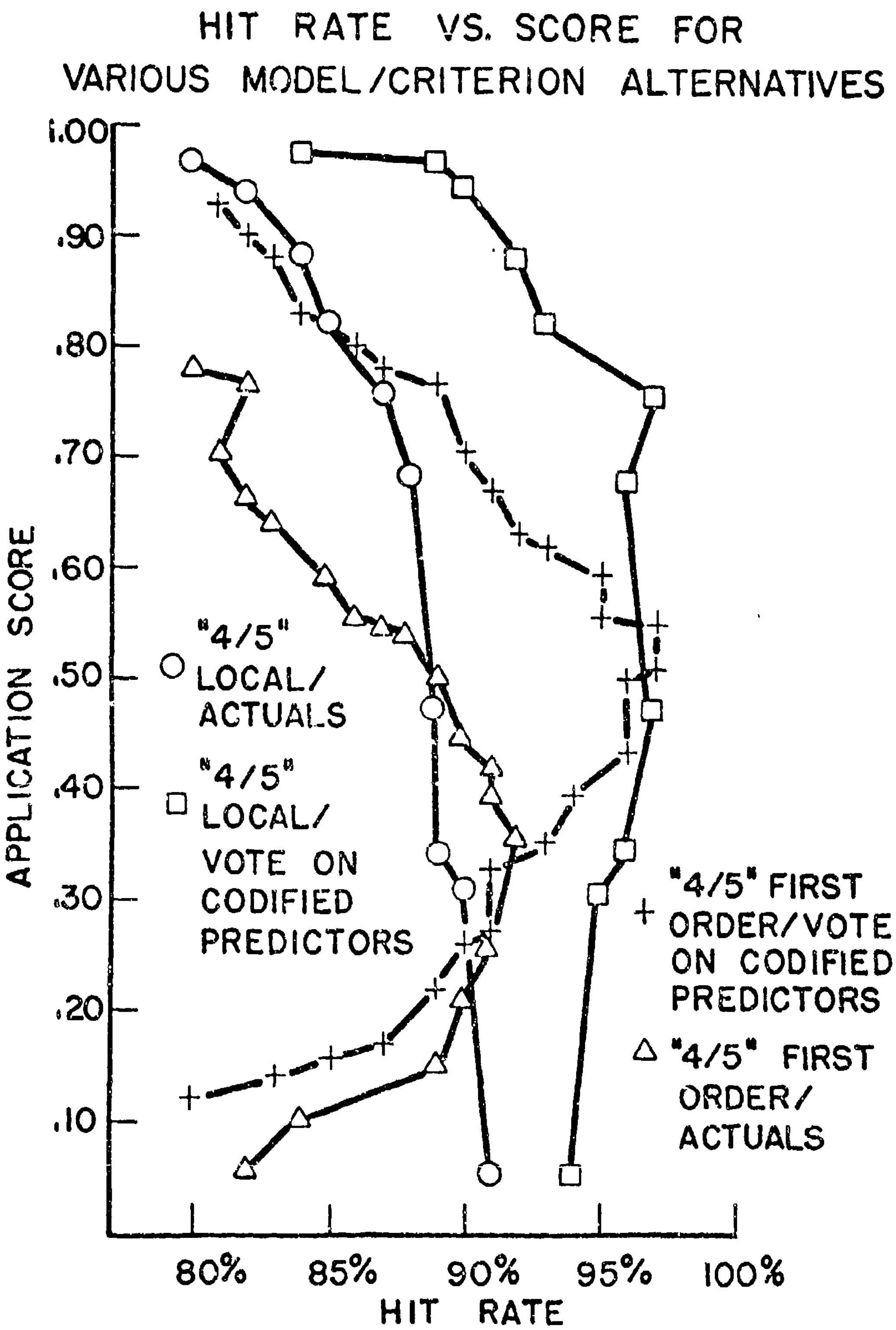

HIT RATE VS. SCORE FOR
VARIOUS MODEL/CRITERION ALTERNATIVES
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
APPLICATION SCORE
"4/5" LOCAL/ ACTUALS
"4/5" LOCAL/ VOTE ON CODIFIED PREDICTORS
"4/5" FIRST ORDER/VOTE ON CODIFIED PREDICTORS
"4/5" FIRST ORDER/ ACTUALS
80%
85%
90%
95%
100%
HIT RATE

FIGURE 6-26

Section VII.

Implications for Implementation:

The results of this modeling effort indicates that there are several applications of the Policy Capturing technique that would be beneficial to the installment loan decision process.

Training:

The discussion and interaction that the process stimulated among the loan officers was felt to be very beneficial. In a post project interview, there was general agreement among the loan officers that the models helped them to understand how their colleagues thought and to illuminate some of the more subtle aspects of their own policies. One specific example of this improved communication was the definition and clarification of the concept of Financial Reference. A second example was best expressed by one of the junior loan officers, who said:

> I really did not know how the three senior loan officers evaluated the applications, other than on past credit, until they discussed their 'trees'.

It would appear that having a new employee make decisions on a standard set of loan applications and then capturing his policy to compare with those the existing group of loan officers would be a simple and effective method of evaluating his compatibility. An example of this was encountered when the transferrability effort was pursued. The policy of a new employee with considerable past experience as manager of another loan company was captured.

The results reflected that he would have given some 38 percent of the applicants with "Poor" Credit Rating loans, while the corresponding success rate for the "4/5" vote was 0 percent. Further, his understanding of the newly defined predictor of Financial Reference and his concern with Bank Account were not commensurate with the attitudes of the other loan officers. The policy capturing exercise made early identification of these differences possible.

Mechanization of Prescoring or "Surrogate" Judgment:

The performance of the "with credit" models appears to be sufficient to allow their use either as an aid to the loan officer, or to automatically approve loans. Although Dawes and Diller's numerical criterion for indicating "bootstrapping" will occur does not appear to be directly applicable in the case of binary decisions, the comparison of the hit rates of the models (See Table 6-15.) with the reliability of the judges (See Table 6-3.) would indicate that any of the models would provide some aid.

Procedures for implementing the models could be manual or computerized. Any of the equations previously presented could be accessed through a computer terminal or stored on a programmable desk calculator to provide the application scor . Figures 6-27 and 6-28 present the risk curves that could be used by the loan officer as he considered an application. On those applications that do not contain significant extenuating circumstances, a fixed cut score could be used for automatic approval or denial. Figure 6-28 indicates that by evaluating applications with the local "4/5" voting model, using a rule of granting loans which scored over .06 and denying loans that scored less than .04 would represent minimum risk and yet leave a very small

RISK CURVES FOR "4/5" POLICY (FIRST-ORDER) WITH CREDIT RATING

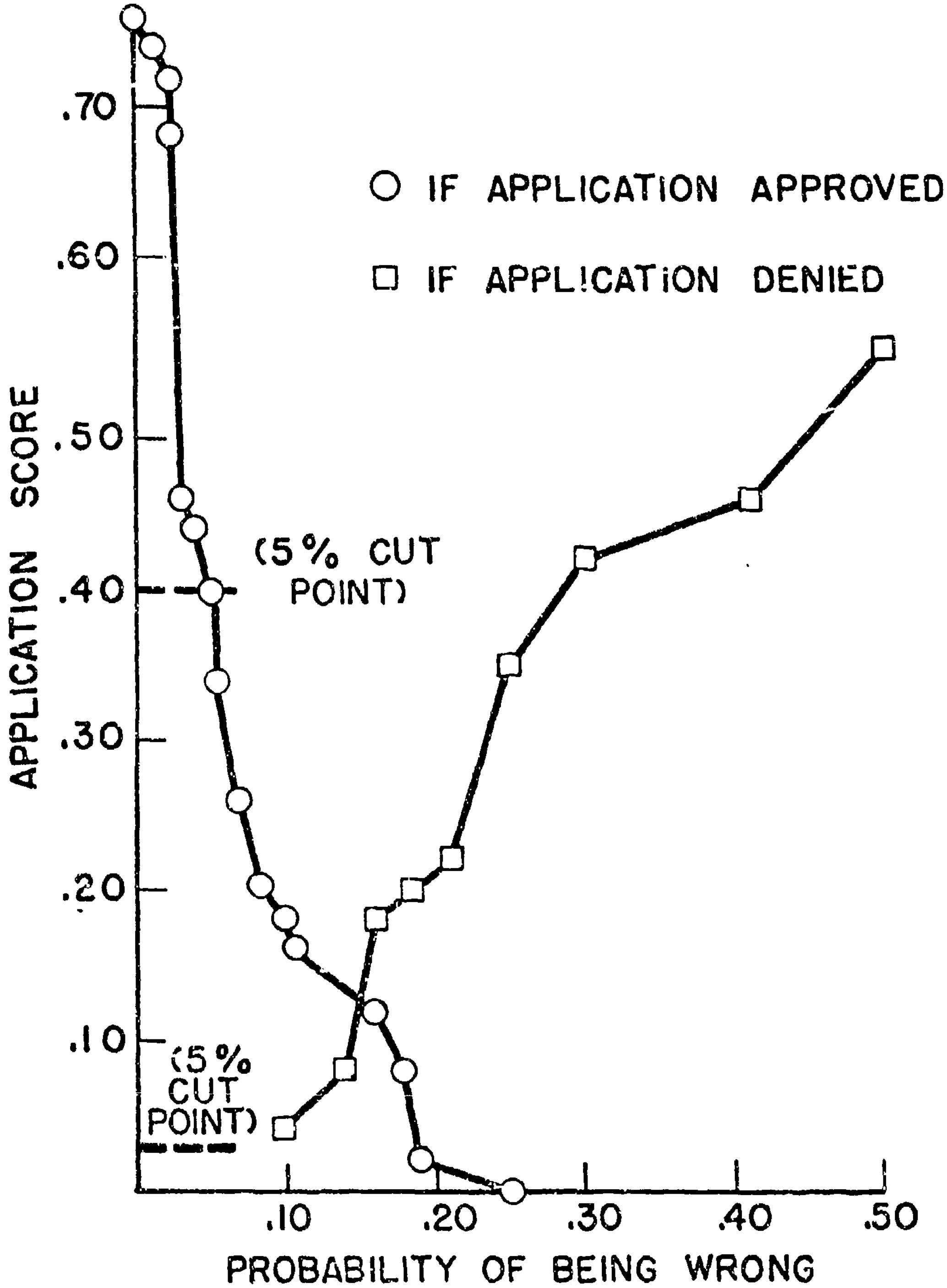


FIGURE 6-27

RISK CURVES FOR "4/5" POLICY (LOCAL MODEL) WITH CREDIT RATING

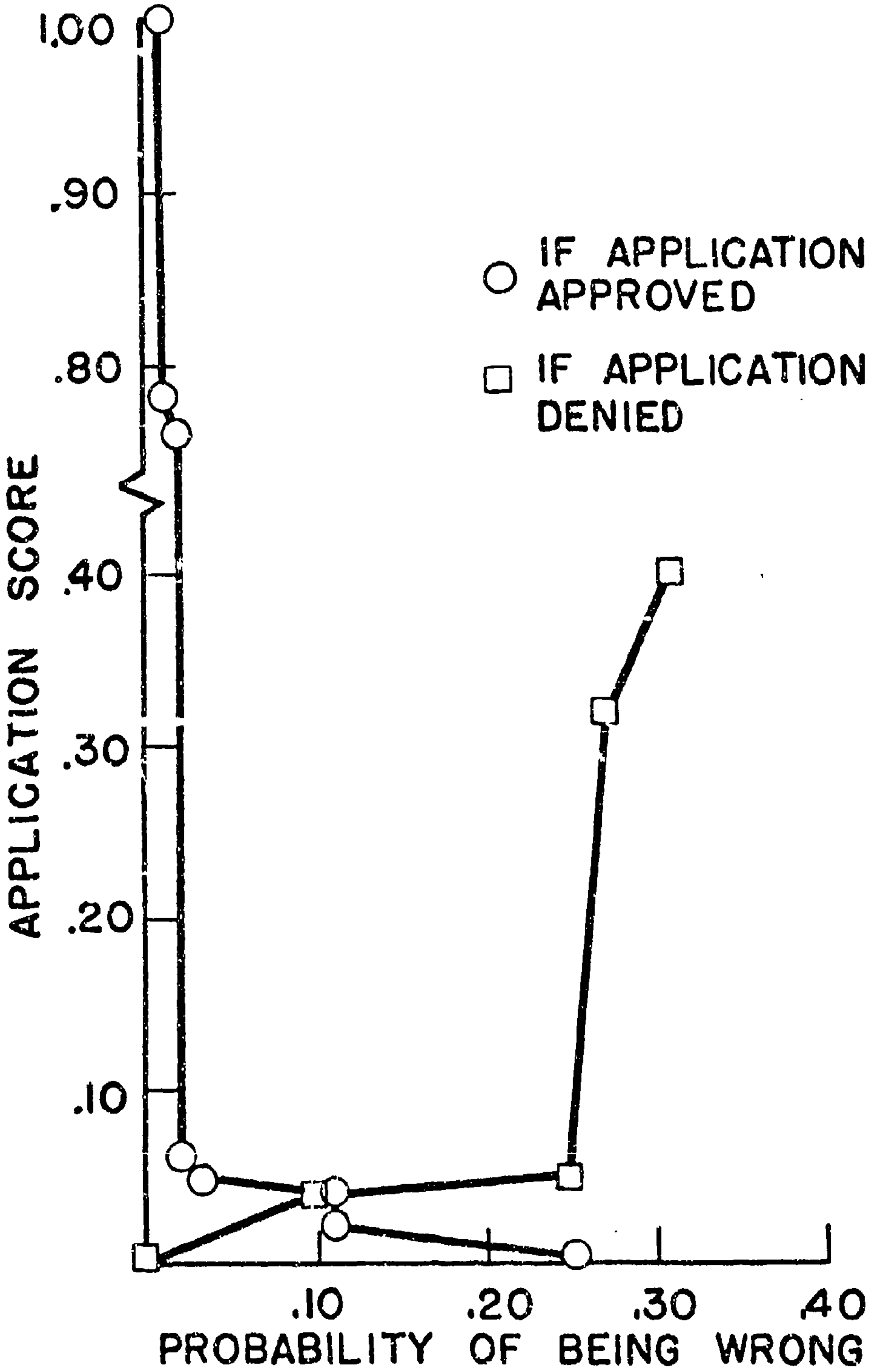


FIGURE 6-28

range of ambiquity.

One ancillary advantage of using the policy models is that the loan officer could get a consensus evaluation from his colleagues by using their models along with his own.

"Tentative Approval" During Non-banking Hours:

The performance of the "tentative approval" models indicates that a definite improvement in the level of service the lender can provide to the dealer could accrue. The motivation for use of these models lies in the speed advantages that the use of the models provides. Figure 6-29 shows the distribution of times that the cross-validation applications spent in the "system" and the corresponding times they spent in the "credit system". The difference in the average times, 14.3 hours, is largely due to the lack of access the dealer has to the lender during non-banking hours. The tentative approval of applications could have eliminated this waiting time for 48 percent of those "after-banking-hours" applications submitted during the field test. This is particularly applicable to sales made on Friday afternoon or Saturday, but which must wait until Monday for consideration.

Again, either manual or computerized techniques could be used. In the simplest system, a manual of predictor categories with application scores (similar to a telephone book) could be provided to the dealer. This approach would be rather inflexible to policy changes. A more sophisticated system could utilize a programmable desk calculator or a remote computer terminal for real-time evaluation of the application.

The difference between the models for "surrogate" judgment and "tentative approval" lies in the current non-availability of a

DISTRIBUTION OF TIMES APPLICATION SPENT IN "SYSTEM" AND "CREDIT SYSTEM"

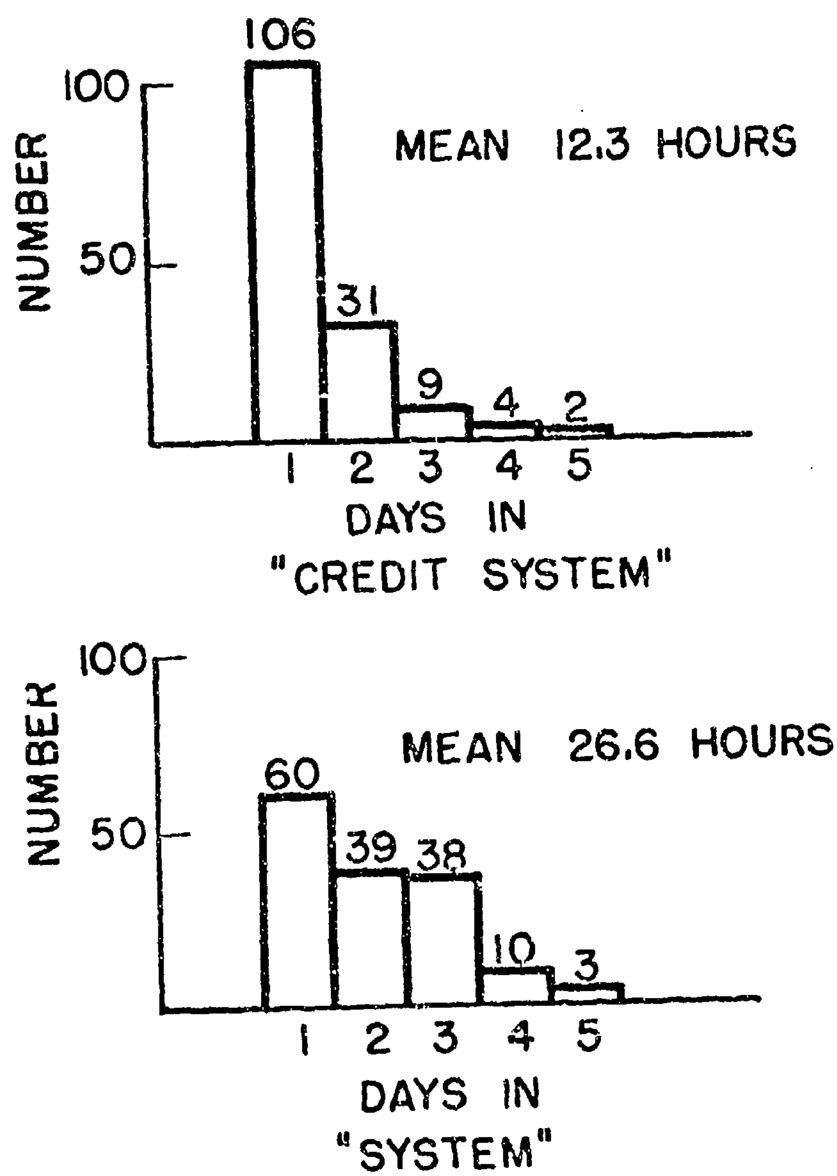


FIGURE 6-29

credit rating at the dealers showroom. With current trends in automation, this condition will likely change with the credit history of an applicant being stored in the credit bureau computer and being accessible electronically via a remote computer terminal. In this situation, much better evaluations could be made and the connotation of "tentative approval" could be removed. To implement a system of this type, the policy model of the lending institution could be stored in a time-sharing computer. At the time the application is made, it could be typed into a remote terminal such as a teletype, the credit rating could be electronically obtained from the credit bureau, and the application score and an appropriate decision could be rendered. An ancillary benefit would be the electronic transmission of the original application to all appropriate parties and the elimination of much of the manual effort currently required from the clerical staffs of the lender and the dealer.

Transferrability of Policy Models:

Two experienced loan officers from other lending institutions were used to assess the potential transferability of the predictor set and the models developed for the loan officers at Lamar Savings. In initial interviews, both of these judges reviewed the predictors and their categorizations, and generally agreed with them. Although each expressed some differences in the exact categorizations of various predictors, these differences were no greater than the differences that had existed among the three senior loan officers from Lamar Savings at the beginning of the project. The most important thing noted was that the only variables suggested by these judges

that were not included in the 16 predictors shown in Exhibit 3, Appendix B, were Sex and Number of Dependents. Both of these were suggested by Judge 7. He noted that his clientele was composed mainly of salaried employees with large families from minority groups and was somewhat more restricted and less affluent than the clientele of Lamar Savings. Neither of the predictors were viewed by Judge 7 as critical omissions.

The AID-Trees for the two judges are shown in Exhibits 14 and 15 of Appendix B. A comparison of their decisions with the vote decisions of Judges 1 through 5 is given in Table 6-17.

TABLE 6-17

COMPARISON OF "4/5" VOTE AND DECISIONS OF JUDGES 6 and 7

Judge	No/No	Yes/No	No/Yes	Yes/Yes	Percent Match
6	61	4	66	93	69
7	98	8	29	89	83.5

Both the AID-Tree and the hit table reflect that Judge 6 apparently has a different policy than that reflected by the majority of the original loan officers. The lower predictive capability of the AID-Tree would suggest that the policy of Judge 6 may not be as well defined or consistent as those of the other judges. The data in the hit table reflects a greatly different policy toward denials. As previously discussed, Judge 6 had just been hired by Lamar Savings from his position as manager of a smaller, more speculative loan company. The high number (66) of approvals by Judge 6 which were denials by the 4/5 criterion would indicate that Judge 6 was

accustomed to assuming more risk by granting applicants with lower Credit Rating loans. His results represent a fortuitous demonstration of the potential use of the Policy Capturing technique to uncover differences in philosophy and background of new employees.

The AID-Tree and hit table for Judge 7 reflect that his policy is much closer to the "4/5" vote of the original judges, but there are still some differences, primarily in the importance of Bank Account and Financial References. These differences are probably reflective of the different clientele since Judge 7's policy is better explained by variables such as Residential Category, Time at Residence, and Income, than it is by variables such as Financial References and Bank Account. It is important to note that the predictive capability of the AID-Tree for Judge 7 is comparable with those of the other five judges. This indicates that the omission of Sex and Number of Dependents as predictors does not greatly degrade the model and confirms his assertion that they are not critical to his policy.

Based on this rather limited insight into transferability, the practice of using the weights of a policy model, derived for one environment, in another environment does not appear to be the best approach. However, the redefinition of these weights by using a standard set of predictor cases does appear feasible.

Section VIII

Summary:

This chapter has presented the major results of the effort to model the policy of installment loan officers under two conditions--with and without access to a Credit Rating on the applicant. Results of both local and first-order models suggest that they could be effectively employed to aid the loan officers in making more efficient and timely decisions. The use of the AID-Tree was found to be a particularly good catalyst to interjudge communication, particularly when the judges are not accustomed to conceptualizing decision processes in terms of algebraic or statistical models.

The results suggest that the local models are more idiosyncratic than the first-order models and are sensitive to policy changes that affect only certain portions of the decision space. However, this sensitivity could be reduced by using a larger set of data to define the local models. The local models do provide a better "average" fit over the entire space.

Models that were able to predict binary decisions, using codified predictors, with 97 percent accuracy were developed. In predicting decisions made with the natural language applications, 92 percent accuracy was attained. In either case, these levels of predictability are better than the 90 percent reliability of the judges. When compared to the 75 percent accuracy that would result if every applicant during July 72 had been given a loan, the improvement in accuracy obtainable with the policy models ranges between 17 and 22 percent. Whereas giving everyone who applies a loan is clearly unwise

in the viewpoint of the loan officer, this latter comparison is not really too informative. The actual economic advantage of using the policy models can only be evaluated after they are used in a decision making role and appropriate cost and profit data has been collected. Comparison of such with the costs and profits for the current mode of operation will then reflect the economic viability of using the models.

The feasibility of identifying the top 30 percent of the applications in the absence of the primary predictor was demonstrated.

Finally, evidence of the utility of the Policy Capturing technique to assess interjudge differences and evaluate the policy of new employees was obtained.

CHAPTER VII: CONCLUSIONS, COMPARISONS, AND RECOMMENDATIONS

Viability of Applying Policy Capturing to Operational Decisions:

The primary goal of this research was the adaptation and demonstration of Policy Capturing in operational environments. Toward this end, the analysis was performed entirely with data taken from an operational environment and the criteria for success were defined relative to the needs and conditions within that environment.

The ultimate product sought was a technique that could be used by the Operations Research practitioner to define, evaluate, compare, and predict the judgments of decision makers in the business, academic, and administrative fields. In order for a modeling technique to be viable in these environments, it must be reasonably straightforward to use and compatible with the stimulus data (in amount and format) as it exists in the environment being modeled. It must produce models that are accurate and understandable.

The results embodied in Table 6-15, Figure 6-23, and Figure 6-28, and the remarks of the junior loan officer, (page 180) are probably the best evidence that can be offered for the utility and viability of Policy Capturing in the operationally valid environment that was studied in this research. They reflect that policy models were obtained which could predict the decisions of installment loan officers with 92 percent accuracy when all of the important predictors are available. Models were also produced which could meet the criterion of identifying the top 25-30 percent of the loan applications even when the most important predictor was not available. The

remarks of the junior loan officer, and the continued participation of the already burdened staff of the installment credit department attest to the believability and communicability of the model.

Comparison of a judge's policy as expressed by any of the algebraic models presented in Chapter VI, with the "structural image" model as expressed by the corresponding AID-Tree, leaves little doubt as to the benefits of graphically displaying the policy. This is especially important in communicating the concept of Policy Capturing to decision makers who are not mathematically inclined.

Modeling Configural Judgment Processes with Local Models:

As for the conceptualization and modeling of configural decision processes as a series of local hyperplanes, the comparison of the models obtained for the loan officers (see Figures 6-14 through 6-18) would appear to support the superiority of this approach from both the accuracy and the interpretability standpoints. Collateral support for the contention that interaction in judgment processes might best be modeled as a discrete phenomena is provided by the comparison of the results obtained using local model with Valenzi's results for "continuous" models.

The Character of the Model Seeking Technique, AID:

In assessing the robustness of the model seeking procedure embodied in the application of the AID4UT/AIDTRE programs, several pertinent points can be made. First, categorical variables are used; this is the most general of all possible representations of the individual

predictors. Second, there are nc underlying restrictions or assumptions imposed in the model seeking phase of the process. Third, the procedure does not require designed experiments or independent predictors. Finally, the capacity of the computer program is large enough to handle up to 80 possible predictors and does not require an "a priori" culling of the variables, or hypothesis of the interactions.

On the other hand, the procedure does not produce a completely definitive answer, but rather, requires the application of considerable judgment on the part of the analyst in the interpretation of the output data. In this sense, the technique merely provides the analyst with the capability of choosing where he wants to inject his wisdom and experience in the modeling process. With more traditional analysis techniques, he makes the decision by choosing the technique and implicitly imposing the appropriate assumptions. In the use of AID4UT/ AIDTRE and local models, he makes his decisions in terms of explicit interpretations of which subspaces and local models to use.

Comparison of Results of this Research with Those of Past Research:

There are some interesting comparisons that can be made between the results of this research and the conclusions ot past investigators. These comparisons generally reflect that the policy models developed for the loan officers are in harmony with the results and expectations of the clincial psychologists.

In reference to past attempts to define configural models, the results of this research reflect an improvement that is chiefly attributable tc the use of the AID technique for determining where to "fish" (Green, 1968) for configural relationships and the modeling of

these interactions as discrete phenomena.

Other specific comparisons include:

Power of First-Order Models, vis-a-vis, Configural Models. Hoffman (1960) and Yntema and Torgerson (1961) demonstrated the power of first-order paramorphics in modeling judgment processes. The results of this study reflect the same situation. Purely on the basis of statistical significance, there is probably no difference between the performance level of the various types of models investigated in this research. However, given the choice of two equally good models (in terms of either explained variation or hit rate), the model that appears to be more reflective of the underlying process would be the better choice since it would be more intuitively satisfying to the judge.

Relative Shrinkage of Configural Models Due to "Overfit". Goldberg (1968) found that configural models of judgment processes tend to be far less stable than their corresponding first-order models. Ward (1954) indicated that configural models generally tended to "overfit" the idiosyncracies of the data, as indicated by much greater shrinkage in the cross-validation coefficient. From these findings, one could infer that if the configural model held up as well as the first-order model upon cross-validation, some evidence of legitimate configural effects exists. The shrinkage for the first-order models of the loan officers averaged 8.7 percent while the shrinkage for their local models averaged 9.7 percent. This difference of only 1 percent would indicate that the "overfit" phenomena is minimal. It would offer evidence that the configurality is real, especially since the cross-validation rates themselves were uniformly higher for the local models (see Table 4-4).

Diminished Configurality in Group Policies. Slovic (1968) noted that models for group judgments tended to be less configural than those for

individuals. This same situation is reflected by comparison of the differences in predicted capability between the first-order models and local models for the "4/5" vote policy with the corresponding differences for the individual loan officers. (See Table 6-15.)

AID-Tree Representations of Einhorn's Non-compensatory Models. The lexicographic model suggested by Einhorn (1970) is conceptually similar to the process depicted by the structure of an AID-Tree. His conjunctive model, in which strong attributes are most important in determining the total score, is analogous to the substitute advantage model of Sonquist (1964) which results in an upper terminating trunk-twig structure in AID-Trees. Similarly his disjunctive model correlates to the substitute disadvantage model as reflected in the lower terminating trunk-twig structure.

Recommendations for Further Research:

This research has served to demonstrate the potential utility of applying Policy Capturing to a practical decision process, but it does not represent a final answer to the applicability of the concept. The project has probably generated more ideas and questions than it has resolved. These new questions can provide the basis for further research projects in both the areas of 1) development of new applications, and 2) improvement of the current modeling techniques and capabilities. Specific projects recommended by this investigator are:

Implementation of Policy Models as Decision Aids for Loan Officers: The models presented in this dissertation could be implemented on

either a programmable desk calculator, or a remote teletype hookup to a time-sharing computer. The primary tasks of such a project would include:

1) Development of the computer code or calculator routine to implement the appropriate models and evaluate incoming applications.
2) Redesign of the loan application form to facilitate collection of the appropriate categorical data.
3) Training of the clerical staff for proper evaluation and coding of the predictor data.
4) Collection and evaluation of data to evaluate actual performance of the policy models in terms of the "ex-post" quality of the approvals and the lost revenue of the denials.
5) Expansion and modification of the current models to allow implementation for other appliance dealers and the development of models for other categories of dealer contracts.

Sequential Generation of Stimulus Data: At the beginning of this research, the incremental generation and consideration of single cases of stimulus data was envisioned as a possible method of constructing policy models with a minimal set of data. At this point it is still unclear how significant changes and trends in the policy could be detected by addition of a single new decision at each stage of the Policy Capturing process. However, a modified version of this concept of incremental data generation does appear to have merit when used with the AID4UT/AIDTRE computer programs in the definition of subspaces and local models.

In development of a policy model, initial analysis should be performed with a sample of actual stimulus data of at least 10 cases per predictor. The identification of homogeneous subgroups would eliminate subspaces within the predictor space which do not require further investigation. In the non-homogeneous groups, past experience indicates that there are often too few data units to determine a stable model. This problem could be solved by generation of an incremental set of synthetic stimulus data which falls entirely within the non-homogeneous subspace. In this manner, the data set would be augmented in those subspaces of the predictor set where the policy of the judge is most complicated and in which there is a requirement for more predictors and more complicated local models.

The task of incrementally generating realistic stimulus data within selected subspaces could be approached by using a synthetic data generation routine such as the one developed by R. F. Fallis of the Ohio State University Psychology Department. This routine was used by Naylor and Wherry (1964) to demonstrate the feasibility of capturing policies with synthetic data. The correlation matrix obtained from the original set of actual decision data would be used by the data generator to assure realistic data was generated within the subspaces under consideration.

A specific research project in this area would include the following tasks:

1) Acquisition and modification, or development, of a synthetic data generation routine.

2) Identification of an appropriate decision process and application of the proposed procedure for defining local models. For example, synthetic data for the non-homogeneous subspaces of the loan officer's AID-Trees might be generated and new models generated.

3) Evaluation of the change in the model quality resulting from the synthetic augmentation procedure.

Estimation of Time to Repair Electronic Equipment: During the past decade, the process of estimating the mean time to repair (MTTR) of various electronic equipment has become of considerable concern to design, reliability, and maintainability engineers. The need for this data is growing increasingly important for use in optimization models for modularizing equipments and determining repair strategies. The problem lies in the fact that the MTTR data is usually needed in the early stages of equipment development, but is not available until after the equipment has been produced and become operational. Only after the equipment has failed and been repaired can observational data be gathered and analyzed. Even if this process were efficient, which it is not, the resultant data is obsolete by the time it is available. The past procedure for obtaining MTTR data needed in the design process has been to infer values for new equipments, based on experience with similar equipments. This process has not been as accurate as is required for use in sophisticated optimization and design routines.

A recent study was sponsored by the U.S. Air Force Materials Laboratory, WPAFB, Ohio, in which the subjective estimates of MTTR values for 23 various pieces of electronic equipments were obtained from experienced maintenance technicians. Although this study had some of the aspects of a Policy Capturing effort, it did not model the technicians judgments as a function of the basic characteristics of the equipment. Instead, estimates for each individual piece of equipment were collected and compared with actual maintenance experience for that piece of equipment. The results indicated a .86 correlation between the estimates of the repair times and the actual repair times (Smith, 1971). This reflects that technicians can make valid estimates of repair times on particular equipments, but it does not improve the situation relative to predicting the repair times for new and different equipments.

It would appear that this situation offers a natural area for the application of policy models. A project in which the estimated repair times for each equipment were modeled as a function of the basic characteristics of the equipment instead of specific equipment's names or stock numbers might provide the model that could be used for predicting MTTR values for new equipments. The project would consist of modeling the MTTR data in terms of parameters which were common among equipments and building models with these parameters as predictors. Specific tasks to be performed in such a project would include:

1) Determination of the basic characteristics of a class of equipments such as function, size, shape, type of assembly, accessibility, type of failure mode, repair procedure, and required test equipment.

2) Acquisition of existing judgmental data from the appropriate agency and description of the individual equipments in terms of these basic characteristics, or generation of new judgment data on new equipments.

3) Definition of models for estimating MTTR values and comparison of the predicted values for similar equipments, not used in the modeling phase, with actual maintenance data that is available in such publications as the FARADA Handbooks.

<u>Prediction of Successful Rehabilitation of Military Prisoners</u>: A project that could be of direct benefit in further evaluating the potential of the local modeling technique proposed in this dissertation is suggested by a study performed by Smith, Gott, and Bottenberg (1967). In that study, various trait and background data was collected on U. S. Air Force prisoners and first-order models for predicting successful return to duty were developed. They achieved predictive efficiencies of 77.4 percent and suggested that better models could probably be obtained with further analysis. Their primary objective was the demonstration of feasibility of the Policy Capturing approach.

During the course of the current research, the need for such a predictive system surfaced as a result of discussions with the personnel from the U. S. Army Correctional Training Facility (CTF), Fort Riley, Kansas. In their case, accurate prediction of potential failures and successes of retrainees could be used to modify the length of stay and the course of instruction required for individual retrainees. Such predictions would conceivably be made on the basis of instructor evaluations obtained in the early part of the rehabilitation program.

This project could also provide insight into the proposed

use of Policy Capturing techniques to evaluate differences in policy between similar governmental agencies. The experience of interacting with two separate governmental agencies would undoubtedly draw upon the managerial talents of the researcher as well as his technical talents and would be a particularly appropriate project for an operation researcher or management scientist.

Specific tasks of this project would include:

1) Acquisition of the appropriate data used by Smith, Gott, and Bottenberg. This data is still on file at the Personnel Research Laboratory, AFHRL, Lackland, AFB, Texas.
2) Analysis of data with the AID4UT/AIDTRE programs and the development of local models.
3) Comparison of the local models with these first-order models obtained in the previous study by Smith, Gott, and Bottenberg (1967).
4) Acquisition and modeling of judgment and actual data from the CTF, Fort Riley, Kansas.
5) Comparison of the data and models on an inter-agency basis.

Final Comments:

The ultimate goal of this research effort was the "demonstration of the potential benefits and the viability of applying the Policy Capturing concept in ecologically valid environments". To be totally complete, such a demonstration must include more examples than were reasonably accomplishable during the 16 months spent on this research.

Fortunately, the examples attempted were successful, knowledge was gained, and tools developed which should make further demonstrations of the concept's viability less laborious.

The projects suggested for further research represent areas in which this investigator believes significant results might be produced in a relatively short time. Other applications are somewhat more esoteric and will require considerably more conceptual innovation. Examples of further applications of Policy Capturing are the development of models for the objectives and desires of voters, preference models and consumer policies in marketing, and the modeling of the objective functions for research and development planners.

The results of this experiment have increased the interest in Policy Capturing and the research will undoubtedly continue. However, there is much to be accomplished before Policy Capturing becomes a completely "canned" technique.

Whatever the final outcome of these proposed projects, some benefit will inevitably be realized for these applications of Policy Capturing. As a minimum, the process of trying to quantify subjective human judgment forces the participating decision maker to think about, and verbalize, his policy. In this respect, the perception that benefits could be derived from mathematically modeling decision processes is not new. As early as 1889, Lord Kelvin recognized the potential of quantifying human judgment by stating:

> I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the nature may be.

APPENDICES

APPENDIX A

Definition and Comparison of Mathematical Models

Purpose:

In the literature covering the past research in Judgment Analysis, various mathematical models are discussed but they are seldom defined in explicit mathematical terms. One of the goals of the present research is the introduction of the model seeking technique, AID, as a tool for improving the policy-model hypothesis process and transforming that process into more of a directed search and somewhat less of a "fishing expedition". Since the many subtle aspects of linear modeling theory cannot all be discussed in this appendix, the objective of this appendix is to merely discuss the various models and computational methodologies considered in this and past research and to relate them to the general linear model.

The basic function of any model is the parsimonious description of nature. Models can be classified as being either deterministic or stochastic. In Policy Capturing we are concerned with stochastic models and the use of various statistical techniques to build and evaluate stochastic models. The model itself represents the relationships that are presumed to exist in the population. Any assumptions that are made as adjuncts to the model pertain to the population.

The General Linear Model: (GLM)[1]

[1] The reader's familiarity with vector and matrix notation and terminology will be assumed throughout this discussion.

A general linear model is a model in which

$$Y = X\beta + \varepsilon \quad \text{or} \quad \text{Expected Value } (Y) = \hat{Y} = X\beta$$

where:

$$Y = \begin{vmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{vmatrix}$$

: an 'n x 1' vector of random observations from the population of all possible Y values

$$X = \begin{vmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & \cdot & & \cdots \\ \cdot & \vdots & & \vdots \\ X_{n1} & & & \end{vmatrix}$$

: an 'n x p' matrix of fixed quantities

$$\beta = \begin{vmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \beta_p \end{vmatrix}$$

: a 'p x 1' vector of unknown parameters

$$\epsilon = \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \epsilon_n \end{vmatrix}$$

: an 'n x 1' vector of random errors

The GLM for a Hypothetical Three-Predictor Problem:

One can always formulate a problem as a general linear model which has unique binary predictors for each possible combination of the predictor variables. For example consider a problem in which there is a criterion that is a function of three predictor variables, $Y_{ijk} + f(R_i, S_j, T_k)$, two of which (R, S) are ordinal, and the third (T) being categorical. Further, consider that there are three distinct levels associated with each of these variable and that the data to be analyzed consists of one point[2] having each of the possible combinations of these predictor values. Table B-1 shows the criterion value for each of the 27 possible combinations (patterns) of the predictor variables.

TABLE B-1

	R $i=1$			R $i=2$			R $i=3$		
	T $k=1$	T $k=2$	T $k=3$	T $k=1$	T $k=2$	T $k=3$	T $k=1$	T $k=2$	T $k=3$
S $j=1$	Y_{111}	Y_{112}	Y_{113}	Y_{211}	Y_{212}	Y_{213}	Y_{311}	Y_{312}	Y_{313}
S $j=2$	Y_{121}	Y_{122}	Y_{123}	Y_{221}	Y_{222}	Y_{223}	Y_{321}	Y_{322}	Y_{323}
S $j=3$	Y_{131}	Y_{132}	Y_{133}	Y_{231}	Y_{232}	Y_{233}	Y_{331}	Y_{332}	Y_{333}

The most general linear model that could be written for predicting the value of a data units possessing one of these 27 predictor combinations would be of the form:

---

[2] In the general case, multiple data points would occur for each predictor-combination. In this case, each criterion value would have an additional subscript denoting that it was the one of several in the (i, j, k)'th cell.

$$Y_{ijk} = \hat{\beta}_1 X_{111} + \hat{\beta}_2 X_{112} + \ldots \quad \hat{\beta}_{26} X_{332} + \hat{\beta}_{27} X_{333} + \epsilon_{ijk} \tag{A}$$

where: Y_{ijk} = the criterion value for that data point having the i'th level of predictor R, the j'th level of predictor S, and the k'th level of predictor T. (the criterion value for that point in the (i, j, k)'th cell)

$X_{ijk} = \begin{cases} 1 & \text{if the point is in the (i, j, k)'th cell} \\ 0 & \text{otherwise} \end{cases}$

$\hat{\beta}_n (n=1, 27)$ = the 27 unknown parameters of the model

ϵ_{ijk} = the error in the predicted value of the point in the (i, j, k)'th cell.

In vector notation corresponding to that given in the definition of the GLM, Y would be a vector of 27 criterion values, X would be a 27 by 27 identity matrix, and $\hat{\beta}$ would be a vector of 27 unknown parameters.

Note that the model that we have presented is not very interesting in that it has exactly the same number of parameters as there are data patterns. Further, there is only one point per cell and each data point can be predicted exactly; therefore, the error vector is null. This model is not particularly desirable in terms of the parsimony with which it describes the data but it does serve as an example of a true model; a true model being defined as:

> A model that does not impose any relationships upon the expected values of the criterion variable which do not exist in nature.

Although a model which has one binary predictor per variable pattern is a <u>sufficient</u> condition for a model to be "true", it is not a <u>necessary</u> condition. A more parsimonious "true"

model might be attainable if certain regularities exist in the relationships between the individual predictors and the criterion values. Such relationships represent restrictions on the GLM that can be explicitly imposed on the model, but which are more commonly implicitly "assumed" to exist in the process of model formulation.

Equivalency and the Regression Formulation of the GLM:

Before preceding with the discussion of the restrictions that are imposed on models as an alternative to representing each possible predictor pattern by its own binary vector, the concept of equivalent models will be introduced. Two linear models (M1 and M2) are said to be equivalent if they meet the following conditions:

1) Both models have the same criterion vector: (Y)
2) Both models have the same number of linearly independent predictors: (rank of X_{M1} = rank of X_{M2})
3) Every vector in the first model is a linear combination of the vectors in the second model.

The multiple regression model is, in fact, equivalent to the general linear model. In the regression formulation, the X-matrix is composed of an n-dimensional unit vector and a 'n x (p-1)' matrix of predictor values. The most general formulation of the regression model is the "categorical" or "binary regression" model. The binary regression model results when the elements of the X-matrix only takes on the values of 0 or 1, and the predictor vectors are interpreted as classifications or attributes. In this model, the x_{np} values of 0 or 1 reflect the absence or presence of the p'th attribute in the n'th data case.

The binary regression model that is equivalent to equation (A) for the hypothetical problem is:

$$Y_{ijk} = \beta_0 + \beta_1 X_{111} + \beta_2 X_{112} + \ldots + \beta_{26} X_{332} + \epsilon_{ijk} \qquad \text{(B)}$$

where:

Y_{ijk} = as defined previously

X_{ijk} = as defined previously

and: $\beta_0 = \hat{\beta}_{27}$, $\beta_1 = \hat{\beta}_1 - \beta_0$, $\beta_2 = \hat{\beta}_2 + \beta_0 \ldots$ etc.

In model (B), the GLM of model (A) has been reparameterized and the new parameters have been referenced to the value of Y_{333} as a result of the elimination of the β_{27} coefficient. There are still 27 parameters with which to estimate 27 data points, thus allowing the prediction of the Y_{ijk} values without error.

The regression model given in (B) could be written in terms of the vector notation given in the definition of the GLM. In that case, Y would be the same vector of 27 criterion values, X would be a partitioned matrix of the form

$$X = \left[\, U \,\middle|\, \begin{array}{c} I \\ \hline O \end{array} \right]$$

where:

U = a 27 by 1 unit vector

I = a 26 by 26 identity matrix

O = a 1 by 26 null row vector

Because of the generality of the multiple regression model and its equivalence to the general linear hypothesis, all other linear models or computational techniques can be viewed as special cases of the regression model upon which various restrictions have been imposed. Each of these special computational techniques can be accomplished by applying multiple regression procedures to equivalent regression models that have been obtained by reparameterization of the general linear model. Although the mathematical equivalence of various experimental designs and special computational models to reparameterized regression models has long been recognized, the specific techniques and procedures to effect such reparameterizations have only recently been addressed by such authors as Bottenberg and Ward (1963), Draper and Smith (1966), Jennings (1967), Mendenhall (1968), Weber (1971), and Ward and Jennings (in press).

Restrictions Imposed on the GLM:

As mentioned previously, there are several types of restrictions that are commonly placed on the general version of the regression formulation of the GLM. Searle (1971) has pointed out that any restriction can be viewed from three different viewpoints. These three viewpoints are:

1) The restriction is arbitrarily imposed as a side-condition solely for the purpose of obtaining a unique solution to the parameters.
2) The restriction is naturally generated as a result of the hypothesis being tested and the presumed state-of-nature.

3) The restriction is part of the basic model.

In this sense, restrictions placed on the GLM to achieve parsimony would fall into category 2. The two most common of these restrictions are:

Constant Difference Restrictions: This restriction implies that the difference in the expected value of the criterion value ($\hat{Y}$) is a constant (K) for any unit difference in the value of the predictor variable. In terms of our hypothetical model this would mean that if there were a unit difference between the 1st, 2nd, and 3rd levels of the variable R, the following relationships would hold:

1 $\quad \hat{Y}_{111} - \hat{Y}_{211} = \hat{Y}_{211} - \hat{Y}_{311} = K$

2 $\quad \hat{Y}_{121} - \hat{Y}_{221} = \hat{Y}_{221} - \hat{Y}_{321} = K$

....

9 $\quad \hat{Y}_{133} - \hat{Y}_{233} = \hat{Y}_{233} - \hat{Y}_{333} = K$

This restriction allows the representation of all levels of the R variable by a single predictor and thus reduces the number of predictor variables in the model. By assuming both of the ordinal variables (R&S) conform to the constant difference restriction, the model for the hypothetical problem could be written:

$$\begin{aligned} Y_{ijk} = \; & \beta_0 + \beta_1 R_i + \beta_2 S_j + \beta_3 T_{k=1} + \beta_4 T_{k=2} + \beta_5 R_i S_j \\ & + \beta_6 R_i T_{k=1} + \beta_7 R_i T_{k=2} + \beta_8 R_i T_{k=3} + \beta_9 S_j T_{k=1} \\ & + \beta_{10} S_j T_{k=2} + \beta_{11} S_j T_{k=3} + \beta_{12} R_i S_j T_{k=1} \\ & + \beta_{13} R_i S_j T_{k=2} + \beta_{14} R_i S_j T_{k=3} + \epsilon_{ijk} \qquad \text{(C)} \end{aligned}$$

where: R_i = the actual value of R at the i'th level

S_j = the actual value of S at the j'th level

$T_{k=1} = \begin{cases} 1 & \text{for data units at the 1st level of T} \\ 0 & \text{otherwise} \end{cases}$

$T_{k=2} = \begin{cases} 1 & \text{for data units at the 2nd level of T} \\ 0 & \text{otherwise} \end{cases}$

$T_{k=3} = \begin{cases} 1 & \text{for data units at the 3rd level of T} \\ 0 & \text{otherwise} \end{cases}$

If the assumption were true, then equation (C) would still represent a "true" model.

No-interaction (Additivity) Among the Predictors: This restriction implies that the difference in the expected value of the criterion value as a result of a difference in the value of one predictor is independent of the value of the other predictors; that is:

$$\hat{Y}_{111}-\hat{Y}_{211} = \hat{Y}_{121}-\hat{Y}_{221} = \hat{Y}_{131}-\hat{Y}_{231} \ldots\ldots = \hat{Y}_{133}-\hat{Y}_{233} = K1$$

etc.

If this type restriction is imposed on the model for the hypothetical example, the resulting model would be:

$$Y_{ijk} = \beta_0+\beta_1 R_{i=1}+\beta_2 R_{i=2}+\beta_3 R_{i=3}+\beta_4 S_{j=1}+\beta_5 S_{j=2}+\beta_6 S_{j=3}$$
$$+\beta_7 T_{k=1}+\beta_8 T_{k=2}+ \epsilon_{ijk} \qquad \text{(D)}$$

where: $R_{i=1}$, $S_{j=1}$, $T_{k=1} = \begin{cases} 1 & \text{for data units at the first level of the predictor} \\ 0 & \text{otherwise} \end{cases}$

$$R_{i=2},\ S_{j=2},\ T_{k=2} = \begin{cases} 1 \text{ for data units at the 2nd level of the predictor} \\ 0 \text{ otherwise} \end{cases}$$

$$R_{i=3},\ S_{j=3} = \begin{cases} 1 \text{ for data units at the 3rd level of the predictor} \\ 0 \text{ otherwise} \end{cases}$$

If both the constant difference and additivity restrictions are imposed upon the GLM, the model takes on its most parsimonious form:

$$Y_{ijk} = \beta_0 + \beta_1 R_i + \beta_2 S_j + \beta_3 T_{k=1} + \beta_4 T_{k=2} + \epsilon_{ijk}$$

where: R_i = actual value of R at the i'th level

S_j = actual value of S at the j'th level

$$T_{k=1} = \begin{cases} 1 \text{ for data units at the 1st level of T} \\ 0 \text{ otherwise} \end{cases}$$

$$T_{k=2} = \begin{cases} 1 \text{ for data units at the 2nd level of T} \\ 0 \text{ otherwise} \end{cases}$$

The Estimation of Model Parameters and Evaluation of Model "Truth" from Sample Data:

The hypothetical example that was previously given considered a population that had only 27 data points, and a "true model" could be written very easily. Generally, models are used to describe the relationships in much larger populations of data with

many more predictor patterns. A parsimonious model that is also known to be true may not be readily apparent. In this case, the analyst must rely upon probability and statistical theory to evaluate the hypothesis that any particular model is "true". This involves the evaluation of the parameters of the model itself as well as the statistics associated with the particular analytical technique that is being employed.

In order to estimate the parameters of a model, sample data is taken from the population and analyzed with the appropriate statistical techniques. The model itself specifies the structural relationships among the variables. The sample data is used to infer the values for the parameters[3] of the model. The quality of these inferences is a function of the deviation of the hypothesized model from a "true model" and the degree to which the sample data resembles the population. Neither the degree of "truth" of a model nor the amount of sampling error can be assessed exactly. However, use of appropriate analytical and experimental procedures can obviously help to avoid the imposition of "untrue" relationships and can help to minimize sampling error. In the first case, avoidance of restrictions and assumptions which impose arbitrary relationships on the variables will reduce the chance of imposing erroneous relationships. In the second case, the use of randomization and the selection of samples that are sufficiently large will help to reduce sampling error.

[3] In order to distinguish between the parameters for the population and the estimates of these parameters derived from the sample, the convention of using Greek letters to designate population parameters and the corresponding English letters to designate the sample estimates will be followed.

Given a model and sample, the analyst generally has a choice of which statistical analysis technique he wishes to use. This choice is tempered by the amount and type of data and the particular information that the analyst wants.

The specific statistical analysis techniques to be discussed in this appendix include: Multiple Regression (MR), Analysis of Variance (ANOVA), Multiple Classification Analysis (MCA), and Automatic Interaction Detection (AID). All of these formulations are actually restricted cases of the general linear model, and, in that sense, they might best be viewed simply as alternative analytical techniques, each of which is applied to data fitting a particular experimental design and each of which most conveniently achieves some analytical goal.

The most important difference among these analytical techniques is not in the basic character of the underlying models; rather, it is in the way in which those models are obtained. In this context, the differentiation of the techniques into "model seeking" and "model hypothesis" categories is appropriate. MR, ANOVA, and MCA fall into the category of "model hypothesis" and AID falls into the category of "model seeking".

A "model hypothesis" technique is one in which a model of the structural relationships among the individual variables of the problem is hypothesized on an a priori basis and the analysis is performed to determine the relevance of the individual variables.

A "model seeking" technique is one in which the model is not hypothesized on an a priori basis, but rather, the data is analyzed and displayed in a form so as to allow model hypothesis to proceed. Model seeking techniques are generally heuristic and mathematically

informal, such as in the case of trial and error approaches. The AID technique represents a step toward formalization of the model seeking process based on the analysis of the data with a sequential variance-reducing procedure.

The Least-Squares Computational Process:

The estimation of the parameters for a multiple regression model has generally been confined to the procedures of least-squares in which the sum of the squared residual errors ($ESS = \sum_{i=1}^{n} \epsilon_i^2$) is minimized. This procedure involves the explicit or implicit solution of a set of 'p' simultaneous equations, known as the normal equations, for 'p'[4] parameters (b_i), that are the estimates of the model parameters (β_i).

Standard computational formulae for the parameters are derived and given in many texts such as Draper and Smith (1966). They are:

Given: $$Y = X\beta + \epsilon$$

the least-squares estimator of β is:

$$b = (X'X)^{-1} X'Y$$

[4] Here 'p' refers to the number of parameters and includes the constant 'b_0' and the 'b_i' weight of each of the predictor variables of the problem.

the predicted values of the criterion are:

$$Y = Xb$$

the residuals are:

$$e = Y - Y$$

and the coefficient of determination is:

$$R^2 = \frac{b'X'Y - ((\sum_{i=1}^{n} y_i)^2/n)}{Y'Y - ((\sum_{i=1}^{n} y_i)^2/n)}$$

and the Error Sums of Squares is:

$$ESS = Y'Y - b'X'Y'$$

There are several algorithms for solving these normal equations, each of which has some particular advantages. The most direct of these algorithm involves explicit solution of the above equations as given in their matrix form. This procedure has one major drawback, if the X'X matrix is singular, there exists no inverse and the equations cannot be solved. This situation exists where there are linear dependencies in the set of predictor vectors and the number of parameters in the model (p) is greater than the number of unique patterns in the data set (m). In cases where $p > m$, the X-matrix is said to be of less than full rank and a unique solution for the parameters estimates

is unobtainable. A less direct approach embodies the iterative determination of the parameters and does not require matrix inversion. If the predictors are independent, this procedure will result in the same parameter values as the explicit solution, and if the predictors do contain dependencies, it will produce one of the infinitude of non-unique solutions that corresponds to the minimum ESS. Thus, use of this latter procedure will allow acquisition of the ESS information required for hypothesis testing relative to full and restricted models even when the individual parameters cannot be uniquely determined.

A computer code that accomplishes the first procedure is the STEP 01 routine of The University of Texas Center for Highway Research, and a code that accomplishes the second procedure is the REGREJ routine of the EDSTAT-J Library (Jennings, 1971) at the UT Computation Center.

A Numerical Example for Comparing the Various Analytical Techniques:

In the comparison of the various analytical techniques, a small numerical example is now presented. The data for this example is given in Table B-2. The general linear model corresponding to equation (A) is given in Table B-3. The corresponding data for the binary regression model is given in Table B-4. Table B-5 reflects the data for the model in which the predictors (X_1, X_2, X_3) have been restricted to by the "constant difference" relationship discussed previously. It does not reflect imposition of the "additivity" restriction. Note that the very small change in the error sum of squares (ESS) between the models in Table B-4 and Table B-5 would indicate that the restriction does not substantially detract from the "truth" of the model.

TABLE B-2

DATA FOR NUMERICAL EXAMPLE

	$X_1=1$		$X_1=2$	
	$X_2=1$	$X_2=2$	$X_2=1$	$X_2=2$
$X_3=1$	12.5	20.5	3	9
$X_3=3$	24.5	36.5	-1	9
$X_3=4$	30.5	44.5	-3	9

TABLE B-3

GENERAL LINEAR MODEL FOR NUMERICAL EXAMPLE

$$Y^T = (12.5, 24.5, 30.5, 20.5, 35.5, 44.5, 3, -1, -3, 9, 9, 9)$$

$$X = \begin{vmatrix} 1&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&0&0&0&0&0 \\ 0&0&0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&1 \end{vmatrix}$$

$$b^T = \beta^T = (12.5, 24.5, 30.5, 20.5, 36.5, 44.5, 3, -1, -3, 9, 9, 9)$$

$$\epsilon^T = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$ESS = 0$$

Note that p=12(number of predictors)/n=12(number of data points)

TABLE B-4
BINARY MULTIPLE REGRESSION MODEL FOR NUMERICAL EXAMPLE

$$Y^T = (12.5, 24.5, 30.5, 20.5, 6.5, 44.5, 3, -1, -3, 9, 9, 9)$$

$$X = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$b^T = \beta^T = (9, 3.5, 15.5, 21.5, 11.5, 27.5, 35.5, -6, -10, -12, 0, 0)$$

$$\epsilon^T = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$ESS = 0$$

TABLE B-5

REGRESSION MODEL WITH CONSTANT DIFFERENCE PREDICTORS FOR NUMERICAL EXAMPLE

$$(Y = \beta_0 U + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \epsilon)$$

$$Y^T = (12.5, 24.5, 30.5, 20.5, 36.5, 44.5, 3, -1, -3, 9, 9, 9)$$

$$X = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 3 & 1 & 3 & 3 \\ 1 & 1 & 1 & 4 & 1 & 4 & 4 \\ 1 & 1 & 2 & 1 & 2 & 1 & 2 \\ 1 & 1 & 2 & 3 & 2 & 3 & 6 \\ 1 & 1 & 2 & 4 & 2 & 4 & 8 \\ 1 & 2 & 1 & 1 & 2 & 2 & 1 \\ 1 & 2 & 1 & 3 & 2 & 6 & 3 \\ 1 & 2 & 1 & 4 & 2 & 8 & 4 \\ 1 & 2 & 2 & 1 & 4 & 2 & 2 \\ 1 & 2 & 2 & 3 & 4 & 6 & 6 \\ 1 & 2 & 2 & 4 & 4 & 8 & 8 \end{vmatrix}$$

$$b^T = (1.24, .038, 7.19, 11.79, -1.70, -7.99, 2.12)$$

$$\epsilon^T = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6, \epsilon_7, \epsilon_8, \epsilon_9, \epsilon_{10}, \epsilon_{11}, \epsilon_{12})$$

ESS(Error Sum of Squares)* = .1433

---

* The data for the b^T and ESS were calculated with EDSTAT-J computer program--see Table B-8. The ESS is explained on the following pages.

The Analysis of Variance Technique:

The Analysis of Variance (ANOVA) technique is a computational methodology which directly apportions the total variance within a set of experimental data into that part of the variance that is attributable to differences in the levels of the predictor variables (treatments) and that part that is due to random error within the treatment levels. ANOVA is a model hypothesis technique that is usually directed at the testing of the null hypothesis that the mean responses for each of the treatment levels (μ_i) are equal, i.e.:

$$H: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$$

There is no single model formulation for ANOVA; the model tested in any given analysis is a function of the underlying experimental design of the data and the hypothesis being tested. Since a discussion of the many aspects of experimental design is not within the scope of this appendix, the discussion henceforth will be confined to the simplest of all ANOVA formulations. The case to be discussed is the fixed-effects model for a single factor experiment in a completely randomized design. Excellent references on both Experimental Design and ANOVA are Kempthorne (1952), Hicks (1964), and Mendenhall (1968).

The One-Way ANOVA Formulation:

Since the objective of analysis of variance procedures is to test the hypothesis of the equality of means, the linear model for the analysis is formulated in terms of deviations of the mean response for each of the treatments from the overall mean response

or grand mean. These differences are known as "effects". The model for the experiment under discussion is:

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij} \qquad \text{(E)}$$

where: Y_{ij} : response of the i'th observation at the j'th treatment level

μ : grand mean $= \sum_{i=1}^{n_j} \sum_{j=1}^{p} \frac{Y_{ij}}{n}$

$\alpha_j = \mu - \mu_j$: effect of treatment "j"

ϵ_{ij} : random error of the i'th observation at the j'th treatment level

If the numerical example problem were to be formulated as a one-way ANOVA problem on the variable X_1 (by ignoring the X_2 and X_3 dimensions of the predictor set) there would be 6 data points at each level of X_1. The appropriate vectors for the model are shown in Table B-6.

TABLE B-6

NUMERICAL PROBLEM AS ONE-WAY ANOVA

$Y^T = (12.5, 24.5, 30.5, 20.5, 36.5, 44.5, 3, -1, -3, 9, 9, 9)$

$$X^T = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{vmatrix}$$

$b^T = \mu, \alpha_1, \alpha_2$

$\epsilon^T = (\epsilon_{11}, \epsilon_{21}, \epsilon_{31}, \epsilon_{41}, \epsilon_{51}, \epsilon_{61}, \epsilon_{12}, \epsilon_{22}, \epsilon_{32}, \epsilon_{42}, \epsilon_{52}, \epsilon_{62})$

Note that there are only two unique data patterns with which to estimate the three parameters, μ, α_1 and α_2. Hence $m < p$ and the situation exists in which a unique value for the minimum ESS can be found but unique values of the parameters cannot be found. Whereas the specific purpose of ANOVA is generally the comparison of α_1 and α_2, this is an unacceptable situation. In actuality, the model that was presented in equation (E) is not complete. The additional restriction of

$$\sum_{j=1}^{p} \alpha_j = 0 \tag{F}$$

is also imposed on the model. This restriction is listed by many authors as an assumption and its actual function is not realized by many students. In reference to Searle's (1971) categorization of restrictions, both viewpoints 2 and 3 appear to be proper interpretations and the ANOVA model is actually composed of both equations E and F.

Since the ANOVA technique is directed at testing of hypothesis and ultimately results in the application of statistical significance tests, the statistical assumption that $\epsilon_i \cong NID(0, I\sigma)$ is generally imposed.

<u>Considerations in ANOVA for Multiple Predictor Experimental Designs</u>:

In the application of ANOVA, certain simplifications in interpretation accrue when the total variance in the data set can be divided directly into independent portions. This can be accomplished with simplified mathematics if the experimental design for the data set is factorial with equal cell frequencies. Although the absence

of this computational nicety does not prohibit application of ANOVA, it does diminish the advantage of testing the hypothesis of equality of means via ANOVA. In the latter case, the test of hypothesis by comparisons of the ESS obtained from full and restricted regression models is a much more attractive alternative. Again it is to be emphasized that the relationship between an ANOVA formulation and a specific regression formulation of any problem is dependent upon the hypothesis to be tested and the experimental design for the problem

As an example of the equivalence of the data obtained by the traditional calculational procedures for ANOVA and the use of full and restricted regression models, consider the data for our numerical problem. Table B-7 shows the ANOVA table that resulted from hand calculations with the formulae from Hicks (1964). Table B-8 shows the output from the EDSTAT-J computer program for the full model shown in Table B-5 and a restricted model in which the $\beta_4, \beta_5, \beta_6$, coefficients were restricted to be 0. This restriction tests the hypothesis that the interaction effects are all 0. The full model is:

$$Y = \beta_0 U + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \epsilon^{(1)}$$

the restricted model is: $Y = \beta_0 U + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon^{(2)}$

Note that the difference in ESS for the two models is:

$$ESS_R - ESS_F = 320$$

which is the same as the difference between the Total Sums of Squares and the Main Effect (X_1, X_2, X_3) Sums of Squares shown in Table B-7.

TABLE B-7

ANALYSIS-OF-VARIANCE TABLE

Source	D. F.	S. S.		M. S.
X_1	1	1704		1704
X_2	1	320		320
X_3	2	168		82
$X_1{*}X_2$	1	9		9
$X_1{*}X_3$	2	292	*	146
$X_2{*}X_3$	2	18		9
$X_1{*}X_2{*}X_3$	2	1		.5
Total	11	2513		

*Subtotal = 320

The Multiple Classification Analysis (MCA) Model and Procedure:

The MCA model was developed by the personnel at the Institute for Social Research at the University of Michigan to assess the relative importance of individual predictor variables in the situation where the predictor variables were correlated. It is a specialized categorical or binary regression model in which the elements of the X-matrix represent group membership of data cases within a particular category of the variables. Whereas Ward (1962)

TABLE B-8

MODEL NUMBER 1

.....AID ..... ITERATION TOLERANCE= .0000010000

CRITERION 9

PREDICTORS 1 - 6

VAR.NO.	DELTA	RSQ.	BTRANV	BTRANRB	ERROR MSQ	PROB	ITER
1	-.82351	.67817	.6782	.6782	80.867	.0010	1
6	.42578	.85946	.8595	.8595	39.237	.0078	2
5	-.16320	.88307	.9163	.9508	30.728	.2395	3
1	.33986	.91140	.6364	.4464	27.029	.1484	4
5	-.07960	.92282	.6642	.4780	24.240	.3081	5

RSQ.= .99994296 MULTIPLE R= .99997148 23

ERROR SUM OF SQUARES= .1433
STANDARD ERROR OF ESTIMATE = .1693 BASED ON 7 PREDICTORS.

VAR.NO.	RAW WT.	STD. WT.	ERROR	ENTRY	VALIDITY
1	.0380	.001312	.0000	1	-.8235
2	7.1948	.248603	.0005	5	.3570
3	11.7929	1.016438	.0004	4	.2586
4	-1.7020	-.128170	-.0009	6	-.3290
5	-7.9944	-1.315117	.0003	3	-.3486
6	5.1292	.350263	-.0009	2	.4258

1.2632 UNIT VECTOR WEIGHT

MODEL NUMBER 2

.....AID ..... ITERATION TOLERANCE= .0000010000

CRITERION 9

PREDICTORS 1 - 3

VAR.NO.	DELTA	RSQ.	BTRANV	BTRANRB	ERROR MSQ	PROB	ITER
1	-.82351	.67817	.6782	.6782	80.867	.0010	1
2	.35705	.80566	.8057	.8057	54.259	.0380	2
3	.25857	.87252	.8725	.8725	40.042	.0767	3

RSQ.= .87251683 MULTIPLE R= .93408609 3

ERROR SUM OF SQUARES= 320.3333
STANDARD ERROR OF ESTIMATE = 6.3278 BASED ON 4 PREDICTORS.

VAR.NO.	RAW WT.	STD. WT.	ERROR	ENTRY	VALIDITY
1	-23.8333	-.823514	0.0000	1	-.8235
2	10.3333	.357048	.0000	2	.3570
3	3.0000	.258571	.0000	3	.2586

28.5000 UNIT VECTOR WEIGHT

had pointed out that Hoffman's (1960) measure of Relative Weight was meaningless in the case of correlated predictors, they sought a measure that would reflect an adjustment for intercorrelation among the variables. Toward this end their procedure calculates "adjusted means"[5] for each category of each predictor.

One difference in the MCA formulation than that commonly used in binary variable regression exists relative to the restriction that is used to insure a full rank X-matrix. In the case of a binary regression formulation where the inclusion of the constant term causes the X matrix to be less than full rank, the model is usually reparameterized so as to include only p-1 categories, making the constant term identical to the mean response for the omitted category. This has the impact of referencing all of the other category "effects" to the mean response for the omitted category instead of the grand mean. In the case of MCA, a full rank X-matrix is attained by restricting the weighted sum of the category means to zero, i.e.:

$$\sum_{j=1}^{p} n_j \alpha_j = 0 \tag{G}$$

The algorithm used to solve the parameters of this model follows an iterative procedure known as the "sweepout" method

[5]Adjusted means are estimates of what the mean value for each category would have been if the predictor vectors had been orthogonal. In this sense, MCA adjusts for non-orthogonalities in the data (Andrews, Morgan, Sonquist, 1967).

(Anderson and Bancroft, 1952) wherein each of the coefficients of the normal equations is sequentially adjusted until further adjustments are below some predetermined threshold level. In describing the relationship between MCA and ANOVA, Andrews, Sonquist and Morgan (1967) describe MCA as a computerized version of ANOVA with unequal cell frequencies. The major difference again is in the restriction of the MCA model, equation G, versus equation F of the ANOVA model. It is pointed out that the difference in this restriction does affect the value obtained for the "treatment" effects in the case of unequal cell frequencies, but does not affect the reduction in the total sums of squares. MCA does not calculate "F-test" values but does output the appropriate sums of squares for such calculations.

The MCA output for the numerical example is shown in Table B-9. Sonquist and Morgan's (1967) description of the output parameters is as follows:

a_{ij} : the deviation of the j'th category mean of the i'th predictor, adjusted for the effects of the other predictors

η_i : the correlation coefficient indicating the ability of a predictor, as categorized, to explain variation in the dependent variable

η_i^2 : the proportion of the Total Sums of Squares explainable by a given predictor "i"

β_i : a coefficient analogous to η_i except based on the adjusted means for each category

TABLE B-9

NO. OF DATA UNITS ANALYZED	=	12	SUM OF Y	=	195.000
WEIGHT VARIABLE SUBSCRIPT	=	-0	STANDARD DEVIATION	=	15.114
SUM OF WEIGHTS	=	12.000	SUM OF Y-SQUARE	=	5681.500
DEP. VARIABLE SUBSCRIPT	=	4			
GRAND MEAN	=	16.250	TOTAL SUM OF SQUARES	=	2512.750
			EXPLAINED SUM OF SQUARES	=	2192.417
			RESIDUAL SUM OF SQUARES	=	320.333

PREDICTOR X1

CODE X(1,J)	NO OF CASES	SUM OF WEIGHTS	PER CENT	C L A S S M E A N	DEV. FROM GRAND MEAN	COEFFICIENT A(1,J)	SUM OF Y	S U M OF Y SQUARE
1	6	6.0	50.0	28.167	11.917	11.917	169.000	5419.500
2	6	6.0	50.0	4.333	-11.917	-11.917	26.000	262.000

UNADJUSTED DEVIATION SS =	1704.08333	ETA-SQUARE =	.67817	ETA =	.82351
ADJUSTED DEVIATION SS =	1704.08333	BETA-SQUARE =	.67817	BETA =	.82351

PREDICTOR X2

CODE X(2,J)	NO OF CASES	SUM OF WEIGHTS	PER CENT	C L A S S M E A N	DEV. FROM GRAND MEAN	COEFFICIENT A(2,J)	SUM OF Y	S U M OF Y SQUARE
1	6	6.0	50.0	11.083	-5.167	-5.167	66.500	1705.750
2	6	6.0	50.0	21.417	5.167	5.167	128.500	3975.750

UNADJUSTED DEVIATION SS =	320.33333	ETA-SQUARE =	.12749	ETA =	.35705
ADJUSTED DEVIATION SS =	320.33333	BETA-SQUARE =	.12749	BETA =	.35705

PREDICTOR X3

CODE X(3,J)	NO OF CASES	SUM OF WEIGHTS	PER CENT	C L A S S M E A N	DEV. FROM GRAND MEAN	COEFFICIENT A(3,J)	SUM OF Y	S U M OF Y SQUARE
1	4	4.0	33.3	11.250	-5.000	-5.000	45.000	666.500
3	4	4.0	33.3	17.250	1.000	1.000	69.000	2014.500
4	4	4.0	33.3	20.250	4.000	4.000	81.000	3000.500

UNADJUSTED DEVIATION SS =	168.00000	ETA-SQUARE =	.06686	ETA =	.25857
ADJUSTED DEVIATION SS =	168.00000	BETA-SQUARE =	.06686	BETA =	.25857

*** ADJUSTED FOR DEGREES OF FREEDOM - MULTIPLE R = .89424 MULTIPLE R-SQUARE = .79967

*** UNADJUSTED FOR DEGREES OF FREEDOM - MULTIPLE R = .93409 MULTIPLE R-SQUARE = .87252

β_i^2 : a measure of the ability of a predictor to explain variation in the dependent variable after adjusting for the effects of the other predictors: Note: This is not in terms of percent of variation explained and the sum over all predictors is not constrained to be less than or equal to unity

R^2 : the multiple correlation coefficient which indicates the proportion of variance in the dependent variable explained by all of the predictors together

The net advantage of using MCA instead of using a standard regression routine to do binary regression appears to lie in the mechanical convenience of defining the predictor matrix and in the fact that the deviations of the mean category responses for all categories are referenced to the grand mean as opposed to 'p-1' categories values being referenced to the p'th category mean.

The Automatic Interaction Detection (AID) Technique:

The AID technique fulfills a different role in the modeling process than the previously discussed techniques. AID is a "model seeking" technique and as such does not presume a specific structure to the data. AID partitions the data set via a stepwise one-way analysis-of-variance procedure and results in a model of the form

$$Y_j = \sum_{i=1}^{n} b_i X_{ij} + e_j$$

where: Y_j : response of the j'th case

b_i : the group mean of 'n' mutually exclusive groups

X_i : a binary indicator of group membership

e_j : the random error of the j'th case from its group mean, b_i

The unique feature of this model lies in the implicit application of the restriction:

$$\sum_{i=1}^{n} X_i = 1 \quad : \quad \text{for all } j$$

In terms of Searle's (1971) categorization of restrictions, this restriction is an inherent part of the model and results from the procedure of sequential "de-clustering" the data units until "n" mutually exclusive groups exists. This de-clustering takes the form of a binary split of the data set based upon that division that will result in the greatest reduction in the error sums of squares. Anderberg (1971) aptly characterized AID as a "monothetic divisive cluster analysis technique" in that it defines clusters by performing sequential binary splits on one variable at a time.

It is important to recognize that AID is nothing other than a particular restricted version of the general linear model. The claim by the authors, Sonquist and Morgan (1964) and Sonquist (1970) that the model is different because it does not require "additivity" could better be stated as:

> AID is a specific case of the general linear model in which the assumption of additivity is trivially satisfied due to the restriction of $\sum_i X_i = 1$ for each of the 'j' data units.

Figure B-1 shows the AID-Tree for the numerical example problem and Table B-10 shows the least squares estimates for the equivalent regression model. In order to reparameterize the AID model into an equivalent regression model, the vector for group 7 was omitted. Therefore, the constant term of the regression output (1) reflects the group mean for group 7 and the other coefficients (2, 3, 4, 5) reflect the deviation of group means of groups 8, 9, 4, and 6 from the mean of group 7.[6]

Being a "model seeking" technique, AID functions to analyze and display the data set in such a way that the latent characteristics and predictor relationships of the data set are emphasized. This emphasis is accomplished by the construction of a binary tree, the branches of which reflect the predictors, and categories of those predictors, that explain the most variance in the response variable. The structure of this tree represents a visual presentation of the structural relationships among the predictors. The analysis of such trees and the hypothesis of models there from is discussed extensively in Chapter IV.

The output of the AID technique is a set of mutually exclusive clusters of similar data units which best explains the variance in the data. Other statistical parameters relative to the variance explainable by each variable in each branch of the tree are also output.

---

[6] Note that there is a factor of 10 difference in the numbers appearing on the AID-Tree and those in the corresponding regression output. A restriction to integer numbers in AID required multiplication of the criterion vector of the numerical problem by a factor of 10.

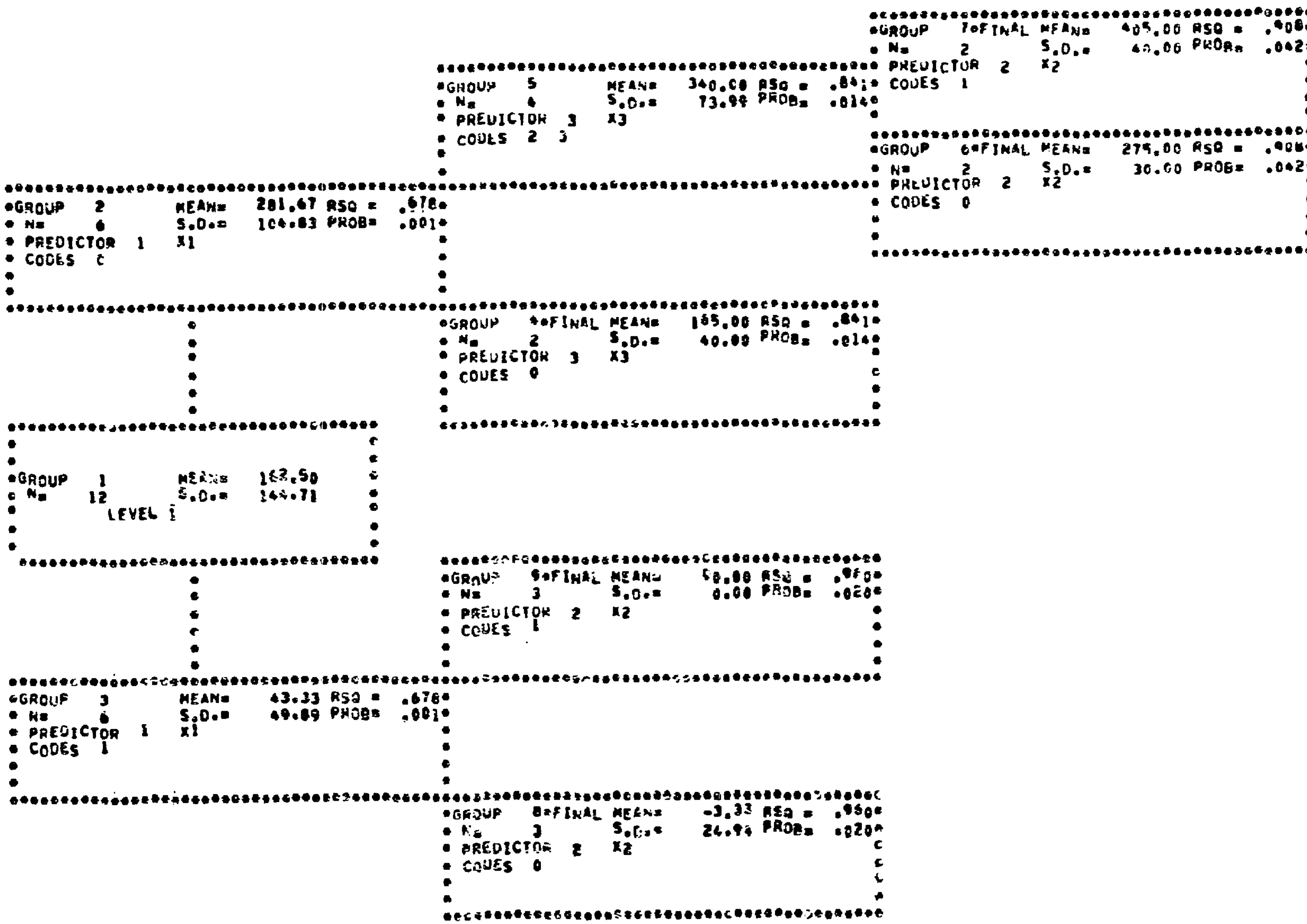

GROUP 1 MEAN= 162.50
N= 12 S.D.= 144.71
LEVEL 1
GROUP 2 MEAN= 281.67 RSQ = .678
N= 6 S.D.= 104.83 PROB= .001
PREDICTOR 1 X1
CODES 0
GROUP 3 MEAN= 43.33 RSQ = .678
N= 6 S.D.= 49.89 PROB= .001
PREDICTOR 1 X1
CODES 1
GROUP 5 MEAN= 340.00 RSQ = .841
N= 4 S.D.= 73.99 PROB= .014
PREDICTOR 3 X3
CODES 2 3
GROUP 4=FINAL MEAN= 165.00 RSQ = .841
N= 2 S.D.= 40.00 PROB= .014
PREDICTOR 3 X3
CODES 0
GROUP 9=FINAL MEAN= 50.00 RSQ = .950
N= 3 S.D.= 0.00 PROB= .020
PREDICTOR 2 X2
CODES 1
GROUP 8=FINAL MEAN= -3.33 RSQ = .950
N= 3 S.D.= 24.94 PROB= .020
PREDICTOR 2 X2
CODES 0
GROUP 7=FINAL MEAN= 405.00 RSQ = .908
N= 2 S.D.= 40.00 PROB= .042
PREDICTOR 2 X2
CODES 1
GROUP 6=FINAL MEAN= 275.00 RSQ = .908
N= 2 S.D.= 30.00 PROB= .042
PREDICTOR 2 X2
CODES 0

FIGURE B-1

TABLE B-10

MODEL NUMBER 1

.....ATD ..... ITERATION TOLERANCE= .0000010000

CRITERION 9

PREDICTORS 1 - 4

VAR.NO.	DELTA	RSQ.	BTRANV	BTRANRB	ERROR MSQ	PROB	ITER
1	-.66165	.43778	.4378	.4378	141.272	.0191	1
2	-.44497	.73018	.5665	.4395	75.333	.0123	2
3	-.28083	.88984	.5643	.3579	34.601	.0093	3
4	-.15379	.94425	.5109	.2764	20.011	.0347	4
1	.08980	.95948	.4514	.2124	14.547	.1489	5

RSQ.= .95993761 MULTIPLE R= .97976406 9

ERROR SUM OF SQUARES= 100.6668
STANDARD ERROR OF ESTIMATE = 3.7922 BASED ON 5 PREDICTORS.

	VAR.NO.	RAW WT.	STD. WT.	ERROR	ENTRY	VALIDITY
(2)	1	-40.8438	-1.222202	.0005	1	-.6616
(3)	2	-31.4815	-.942047	-.0007	2	-.2893
(4)	3	-23.9987	-.618070	-.0000	3	.0077
(5)	4	-13.0054	-.334946	.0002	4	.3477

(1) 43.4987 UNIT VECTOR WEIGHT

The objectives of the application of this analysis technique are different than the objectives of those techniques in which a "true model" is presumed and then statistics are generated for the testing of hypothesis. Consequently, there is really no direct compairson between it and the other techniques. AID is best described as a predessor to the application of these other techniques. However, AID does result in a valid model that is a restricted version of the general linear model. If the mutually exclusive groups into which AID divides the data set are found to be valid within the theoretical context of the data being analyzed, it can be viewed as a binary regression model about which various hypotheses may be tested.

A computer listing and user's description for the AID4UT/AIDTRE program is included as Appendix C of this dissertation.

Appendix B: Exhibits 1-19

This appendix consists of various forms, instructions and computer outputs used in this research and referenced in the body of the dissertation. They are included here as supplementary exhibits.

Exhibit 1
ORIGINAL APPLICATION FORM

Approved by Savings and Loan Commissioner of Texas for use under Consumer Credit Code

LOAN APPLICATION
(Consumer Credit Code Article 3.15 or Chapter 4 or 5)

TO Lamar Savings Association (Name of Association) Date: ______ 19__

I hereby apply for membership and a loan under the Consumer Credit Code for the purposes set out below:

Net Amount Required $______ Number of Months ______

(1) Identification: Name ______

Address ______ (Street) (City) (State) (Zip Code)

At present address ______ years. Telephone ______ Age ______ Single ☐ Married ☐

Previous address ______ years

Name of wife (or husband) ______ (Age) Number of Dependents ______

If Military: Serial No. ______ If Civil Service. Badge No. ______ Soc. Sec. No. ______

Name and Address of Relatives Not Living with Applicant: ______

(2) My Employment or Business is: Employed by ☐ Self-employed ☐ ______ (Name of Firm) Past ______ years.

Business Address ______ Business Telephone ______

Kind of Business ______ My Position ______

Present Salary or Net Income from business $______ per month ☐ per year ☐

Wife's Salary $______ per month ☐ per year ☐ Other Income (Net) $______

Wife's Employer ______ for past ______ years.

Previously Employed by ______ (Name and Address) for ______ years.

(3) I have ☐ received credit ☐ credit applications pending at the following firms:

1. ______ 3. ______
2. ______ 4. ______

(4) My Bank is ______

(5) I have debts outstanding as follows:

Describe Debts	To Whom Indebted (Name)	Date Incurred	Original Amount	Present Balance	Monthly Payments
			$	$	$

(6) The proceeds of this loan will be used:

☐ To improve real property on which one to four family dwelling ☐ is ☐ is not ☐ will be located as follows:

(a) Improvements to be made ______

(b) Property Identification: Lot ______ Block ______
Address ______ Type of Structure ______
My title is by ☐ deed dated ______ Recorded Vol. ______ page ______
☐ contract to purchase dated ______ from ______

(c) Present Liens against property. Amount $______ Held By: ______

☐ purpose: (specify) ______

(7) I understand that the following insurance coverage will be required in connection with this loan and that I have the option of furnishing such insurance either through existing policies owned or controlled by me or procuring and furnishing equivalent coverage through any insurance company authorized to transact business in Texas:

1. ☐ Credit Life on ______ Obligor(s)
2. ☐ Credit Health and Accident on ______ Obligor(s)
3. ☐ Fire and extended coverage
4. ☐ Comprehensive Coverage
5. ☐ Fire, Theft & Additional Coverage
6. ☐ $______ deductible collision
7. ☐ Other:

I will furnish the following coverage:
and request you to provide and charge me for any not furnished by me.

(8) I further understand that the selection of a contractor or dealer, acceptability of materials used and work performed for the funds you advance me as a result of this application is my responsibility and you do not guarantee material, workmanship or inspect work performed for me. Each person signing this application certifies that the above statements made for the purpose of inducing the extension of credit are true and that no unfavorable information bearing on the trustworthiness of the person or persons signing has been omitted and that each statement is applicable to each individual whose name is signed hereto unless otherwise indicated.

READ THE FOREGOING STATEMENT ______

CAREFULLY BEFORE SIGNING. ______

Exhibit 2

ORIGINAL PREDICTOR VARIABLES AND CATEGORIES

Variable	Categories
Age:*	< 23, 23-30, 30-50, 50-63, > 63
Marital Status:	Married, Divorced, Single
Local Family:	No, Yes
Draft Status:*	Eligible, Ineligible
Telephone:	Yes, No
Average Time (Residence and Job)*	< 2, 2-10, > 10
Residence Type:	Rent, Buying, Own, Rent Free
Type Employment:*	Unskilled, Self-employed, Military, Executive, Skilled
Income (Per Month):	< $200, 200-400, 400-500, 500-600, 700-800, 800-1000, > 1000
Checking or Savings:*	Yes, No
Credit Habit (who he owes):*	Good, Bad
Outstanding Debts:*	< 500, 500-1000, 1000-2000, 2000-3000, 3000-4000, > 4000
Loan Amount:	< 150, 150-250, 250-400, 400-600, 600-800, > 800
Loan Term (months):*	6-12, 15-18, 24, 36
Item Type:	Necessity, non-necessity
Equity:	< 10%, 10-20%, > 20%
Past Credit Rating:	Poor, Unevaluated, Medium, Good
Application Rating:	Bad (0-2.4), Good (2.5-5)
Actual Response:	Approved, Disapproved

*Variables that were later redefined or re-categorized.

Exhibit 3
REVISED PREDICTOR VARIABLES AND CATEGORIES

	Variable	Categories
1.	Age:	(1) <21, (2) 22-24, (3) 25-30, (4) 31-40, (5) > 40
2.	Marital Status:	(1) Married, (2) Single, (3) Divorced
3.	Local Family:	(1) No, (2) Yes
4.	Telephone:	(1) No, (2) Yes
5.	Time on Job:	(1) <2, (2) 2-5, (3) >5
6.	Time in Residence:	(1) < 2, (2) 2-5, (3) > 5
7.	Residence Category:	(1) Rent free, (2) Renting, (3) Buying, (4) Own
8.	Type Employment:	(1) Misc., (2) Unskilled, (3) Skilled, (4) Executive
9.	Income (monthly):	(1) < $400, (2) $400-$600, (3) $600-$900, (4) > $900
10.	Bank Account:	(1) None, (2) Savings, (3) Checking, (4) Other
11.	Financial Reference:	(1) Other, (2) Minor, (3) Major
12.	Loan Amount:	(1) < $300, (2) $300-$600, (3) $600-$900, (4) > $900
13.	Loan Term:	(1) 1 year, (2) 2 years, (3) 3 years
14.	Purchase Type:	(1) Non-necessity, (2) necessity
15.	Equity:	(1) < 10%, (2) 10-20%, (3) > 20%
16.	Credit Rating:	(1) Poor, (2) Unrated, (3) Medium, (4) Good

Exhibit 4 -- DATA FCR 224 SAMPLE APPLICATIONS

1	1	1	1	1	2	1	1	2	3	3	_	3	1	1	2	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	101000
2	1	1	1	2	2	1	1	2	4	1	1	2	2	1	1	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
3	1	2	1	2	2	3	3	2	1	1	1	3	1	1	1	2	1	0	1	1	1	1	0	0	0	0	1	1-0	0	0	1	1	0	000010
4	1	2	1	2	2	2	2	2	2	2	1	2	3	3	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
5	1	5	1	2	2	1	3	3	3	2	1	3	1	2	2	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	1	100000
7	1	3	1	2	1	1	2	3	3	1	1	1	1	2	2	1	1	0	0	0	0	1	0	0	0	0	0	1-0	0	0	0	0	0	000000
8	1	3	1	1	2	3	1	2	1	2	3	2	2	2	1	1	3	0	0	0	1	0	0	0	1	1	0	1-0	0	0	0	0	0	011110
9	1	3	1	2	2	1	1	2	3	3	3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
10	1	4	1	1	2	1	1	2	4	3	1	3	3	3	1	2	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
11	1	3	1	1	2	1	1	3	2	3	4	2	2	2	2	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	1	000010
12	1	2	1	2	2	2	2	2	2	2	1	2	2	1	1	2	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	0	101110
13	1	4	1	1	2	3	2	2	2	3	1	2	1	1	1	3	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	100000
14	1	4	1	1	2	3	2	2	3	3	3	2	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
15	1	4	1	2	2	3	3	2	2	1	3	1	1	2	1	3	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
16	1	3	1	1	2	2	2	1	3	4	4	2	3	3	2	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
17	1	1	1	2	2	1	1	2	2	3	1	3	2	3	1	1	1	1	1	0	1	0	0	0	0	0	0	1-0	0	0	0	1	0	000000
18	1	4	1	2	2	2	3	2	3	4	3	3	2	2	1	1	1	1	1	1	1	1	0	0	0	0	0	0-0	0	0	1	1	0	000000
19	1	5	2	1	2	2	2	3	3	3	3	2	1	2	1	1	1	0	0	1	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
20	1	2	1	1	2	1	1	2	3	2	1	2	2	2	1	1	1	1	1	0	1	0	0	0	0	0	0	1-0	0	0	0	1	0	000000
21	1	2	3	1	1	1	2	1	1	1	3	1	1	2	1	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
22	1	2	1	1	2	1	2	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
23	1	3	1	2	2	1	2	3	1	3	2	2	3	3	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
24	1	4	1	2	2	1	1	2	2	4	1	1	1	1	1	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
26	1	3	1	2	2	2	3	3	3	3	3	3	2	2	2	1	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	113101
27	1	2	1	2	2	1	3	2	2	3	1	3	2	3	2	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	101111
28	1	4	1	2	2	1	1	1	4	3	3	3	2	3	1	3	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	100000
30	1	2	1	1	2	2	2	2	2	1	3	3	1	2	1	1	2	0	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	001111
31	1	5	1	1	2	3	3	3	2	3	3	3	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0-0	0	0	1	1	1	100010
32	1	4	3	2	2	1	3	2	2	1	1	3	1	1	1	1	2	1	0	1	0	1	0	1	0	1	0	1-0	0	0	0	1	0	101000
34	1	3	1	2	2	2	3	2	4	2	2	3	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0-0	0	0	0	1	0	100000
36	1	2	1	1	1	1	1	2	2	2	3	3	2	3	1	1	3	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100010
38	1	5	3	2	2	1	1	3	3	4	3	3	2	3	2	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	100000
39	1	5	1	2	1	3	3	3	2	1	1	2	2	3	2	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
40	1	2	1	2	2	1	1	2	3	1	1	2	2	3	2	1	2	1	0	1	1	0	1	0	0	1	0	1-0	0	0	0	1	0	000000
41	1	4	1	1	1	3	3	3	4	4	3	1	2	3	1	1	2	0	1	0	0	1	0	1	0	0	1	1-0	0	0	0	0	1	101100
42	1	3	1	1	2	3	1	3	2	2	3	3	2	3	2	1	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	0	111111
43	1	3	1	2	2	1	2	2	3	2	1	2	3	3	1	1	2	1	1	1	1	1	1	0	0	1	1	1-0	0	0	1	1	0	101000
44	1	2	1	1	2	1	1	2	2	4	3	3	3	3	1	2	1	1	1	1	1	0	0	0	0	0	0	1-0	0	0	1	1	0	100000
45	1	5	3	2	2	3	3	1	2	1	1	3	1	1	1	3	2	0	1	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	101011
46	1	1	1	2	1	2	1	2	2	1	1	2	1	2	1	2	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
47	1	4	1	2	2	3	2	2	2	3	1	2	2	3	1	2	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	101100
48	1	1	1	2	2	1	1	3	3	2	3	3	2	3	2	1	3	1	1	1	1	1	0	1	1	1	1	1-0	1	1	1	1	1	101110
49	1	3	3	1	2	2	1	2	4	2	1	1	1	1	1	3	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	100000
51	1	5	1	2	2	1	3	3	2	3	3	1	2	3	1	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
52	1	5	1	1	2	3	2	2	3	4	4	3	2	2	1	1	1	1	1	1	1	1	0	0	0	0	0	0-0	0	0	1	1	0	100000
53	1	2	3	2	1	1	2	3	2	1	1	2	1	2	1	1	2	0	0	0	0	0	0	0	0	1	0	1-0	0	0	0	0	0	100000
55	1	3	3	1	1	1	1	2	2	2	4	3	2	2	1	1	2	0	1	1	1	1	0	0	1	1	0	1-0	0	0	1	1	1	000100
56	1	4	1	1	2	3	2	2	4	3	1	1	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
57	1	4	1	2	2	3	3	3	3	3	3	3	1	2	2	2	1	1	1	1	1	1	0	0	0	1	1	0-0	0	0	1	1	1	100000
59	1	3	1	2	2	3	3	2	4	3	3	3	2	2	2	2	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
60	1	4	1	2	2	2	3	3	3	3	3	3	1	2	2	1	1	1	1	1	1	1	0	0	0	1	1	0-0	0	0	1	1	0	100000
61	1	5	1	2	2	3	1	2	4	4	4	2	2	2	2	2	3	0	0	1	0	0	1	0	1	1	0	1-0	1	0	0	0	1	111101
62	1	3	1	2	2	2	2	3	3	1	1	2	1	2	1	2	1	1	0	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	100000
63	1	2	1	2	1	1	1	2	2	2	3	3	3	3	1	1	2	0	0	0	1	0	1	0	0	0	0	1-0	0	0	0	0	0	001000
67	1	4	2	1	2	1	1	3	3	4	3	3	2	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
69	1	4	1	2	2	3	3	2	3	3	3	1	1	1	1	1	1	0	1	0	0	1	0	0	0	0	0	1-0	0	0	0	0	0	100000
70	1	1	1	2	2	1	1	2	2	2	1	2	2	3	1	1	3	0	0	1	1	0	0	0	1	1	0	1-0	0	0	0	0	0	000010
71	1	1	1	2	2	1	3	2	3	3	3	1	2	3	1	1	4	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	111100
72	1	4	1	2	1	1	1	3	4	4	3	2	1	2	1	1	1	0	1	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
73	1	2	1	2	1	2	2	3	3	4	3	3	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
74	1	2	1	2	2	2	2	3	4	3	1	3	2	3	1	1	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
76	1	4	1	1	1	1	2	2	2	4	1	3	1	1	1	3	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
77	1	2	1	1	1	2	1	2	1	2	1	3	2	3	1	1	1	0	0	1	1	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
78	1	5	1	2	1	3	2	2	2	2	1	2	2	2	2	1	2	0	0	1	0	1	0	0	1	1	1	1-0	0	0	0	0	0	101110
80	1	5	2	2	2	3	3	2	4	1	1	2	1	2	1	1	1	1	1	0	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
81	1	4	1	2	2	2	3	2	2	4	3	3	2	3	1	1	1	1	1	1	1	1	0	0	0	0	0	0-0	0	0	1	1	0	000000
82	1	5	2	1	2	2	2	3	4	2	3	3	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	101001
83	1	5	1	2	2	2	3	3	2	2	3	2	2	2	2	1	1	1	1	1	1	1	0	0	0	1	0	1-0	0	0	1	1	0	100310
84	1	4	1	2	2	3	3	3	2	4	3	3	2	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0-0	0	0	1	1	0	100000
85	1	2	1	1	1	1	1	2	1	2	3	3	1	1	2	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
86	1	3	1	2	2	1	2	3	2	2	1	3	1	2	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
88	1	3	1	1	2	2	2	2	3	3	1	1	1	2	1	1	3	0	1	0	0	0	0	0	1	1	0	1-0	0	0	0	0	1	100100
90	1	3	1	2	1	2	1	2	2	3	2	1	1	2	1	1	1	0	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
92	1	4	1	2	2	1	3	3	2	1	1	2	1	2	2	1	2	1	0	0	1	1	1	1	0	1	1	1-0	1	1	0	0	1	101000
93	1	2	3	1	2	1	3	2	2	1	3	3	1	1	1	1	1	0	1	0	1	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
94	1	4	1	1	2	1	1	1	2	3	3	3	2	3	1	1	1	1	1	1	0	0	0	0	0	0	0	1-0	0	0	0	1	0	000000
95	1	5	1	2	2	3	3	3	4	1	3	3	2	3	1	2	3	1	1	0	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
96	1	3	1	1	2	1	1	1	4	4	3	3	2	2	2	1	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
97	1	2	1	2	1	1	1	1	3	3	1	2	1	2	1	1	2	1	0	1	0	0	0	0	0	1	1	1-0	0	0	0	0	0	000000
98	1	3	1	1	2	3	3	2	2	2	3	2	1	2	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	100000
99	1	4	1	1	1	1	2	2	2	3	1	2	1	1	2	1	3	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	101010
101	1	4	2	2	2	2	3	3	2	1	3	1	2	3	1	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
104	1	3	1	1	2	1	1	2	2	2	3	1	2	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
105	1	1	1	1	1	1	1	2	2	3	2	3	1	2	1	1	1	0	0	1	1	0	0	0	1	0	0	1-0	0	0	0	0	0	000000
106	1	5	1	2	2	1	1	2	3	2	3	3	2	3	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
107	1	3	1	1	2	1	1	2	2	1	3	2	2	3	1	1	1	0	0	1	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
108	1	4	1	2	2	3	2	2	4	3	3	3	1	2	2	1	1	1	1	1	1	1	0	0	0	0	1	0-0	0	0	1	1	0	100000
109	1	2	1	2	2	2	2	2	2	1	1	1	1	2	1	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
110	1	4	1	2	2	2	2	2	2	1	3	1	2	3	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
112	1	3	2	2	2	1	2	1	4	1	1	2	1	2	2	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	101000
115	1	5	1	2	2	3	3	4	2	2	3	1	1	2	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
116	1	2	2	2	1	1	1	2	2	2	1	2	2	1	1	2	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
117	1	2	1	1	2	1	2	2	3	3	3	3	1	2	2	1	1	1	1	1	1	1	0	0	0	0	0	0-0	0	0	1	1	0	000000
118	1	3	1	1	2	1	1	2	2	2	3	2	1	3	2	1	3	1	0	1	1	1	1	1	1	1	0	1-0	1	1	1	1	0	000010
119	1	5	1	1	2	1	1	4	4	4	4	3	2	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	100010
120	1	5	1	1	2	2	2	2	2	4	4	1	2	3	1	2	3	0	0	1	0	1	0	1	1	0	0	1-0	0	0	0	0	1	101100
121	1	4	2	2	2	1	2	2	3	3	3	1	1	3	1	1	1	0	0	0	0	1	0	0	0	0	0	1-0	0	0	0	0	0	000010
122	1	3	1	2	1	2	1	2	3	4	1	2	1	3	1	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	001000
123	1	5	1	2	1	3	3	2	2	1	2	2	2	3	2	1	3	0	0	1	0	0	0	0	0	0	0	1-0	0	0	0	0	1	101111
124	1	5	1	1	1	2	2	2	3	2	1	3	1	1	1	2	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
126	1	5	3	2	2	2	3	3	2	1	3	3	1	2	1	1	1	0	1	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
127	1	5	1	1	2	1	2	2	2	4	3	3	1	1	2	3	1	1	1	1	1	1	0	0	0	0	0	0-0	0	0	1	1	0	100000
131	1	3	1	1	1	1	2	2	3	2	4	2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
132	1	3	2	1	2	1	2	2	2	1	3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
136	1	3	1	1	2	1	3	2	2	3	2	3	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0-0	0	0	1	1	1	100000
137	1	4	1	1	2	2	2	2	3	4	1	2	2	2	1	2	1	0	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	0	000000
140	1	4	3	1	2	2	2	2	2	1	1	1	3	3	1	1	4	0	0	0	0	0	0	1	1	1	0	1-0	1	0	0	0	1	101100
142	1	5	1	1	2	1	3	2	1	1	1	2	2	3	2	1	2	0	1	0	1	0	0	0	0	0	0	1-0	0	0	0	0	1	001000
146	1	5	1	2	2	3	3	2	2	1	3	3	1	2	1	1	1	1	1	0	1	1	0	0	0	0	0	1-0	0	0	1	1	1	100000
149	1	3	1	2	2	2	3	2	3	2	3	3	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0-0	0	0	1	1	0	100000
150	1	4	1	2	2	1	1	2	2	3	1	3	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
153	1	3	2	1	2	1	2	2	2	1	1	2	3	3	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
154	1	2	1	2	1	1	2	2	2	3	1	3	2	3	1	1	1	0	0	1	1	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
156	1	5	1	2	2	3	3	2	2	3	3	1	3	3	1	1	3	0	1	1	0	1	0	0	0	0	1	1-0	0	0	0	1	0	101100
158	1	3	1	2	2	1	2	3	3	3	3	2	2	2	1	1	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
159	1	3	1	2	1	1	2	2	3	1	1	1	1	2	1	1	3	0	0	0	1	0	0	0	0	1	0	1-0	0	0	0	0	0	000000
160	1	5	3	2	2	1	3	3	4	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
163	1	2	1	1	2	1	1	3	2	2	1	2	3	3	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
166	1	5	2	1	2	3	2	2	2	1	1	1	1	2	1	1	3	0	0	0	0	0	0	1	0	1	0	1-0	0	0	0	0	0	101000
167	1	3	1	2	1	1	2	2	2	1	1	2	1	2	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
168	1	2	1	2	1	1	2	3	2	2	1	2	2	2	1	1	1	0	0	1	1	1	0	0	0	0	0	1-0	0	0	0	0	0	003000
169	1	4	1	1	2	1	2	3	2	2	1	3	1	2	1	1	1	1	1	1	1	1	0	0	0	0	0	1-0	0	0	1	1	1	000000
170	1	2	1	1	2	1	1	2	2	3	1	3	2	3	1	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000

171	1	5	1	2	2	1	1	2	2	2	1	3	2	3	1	1	1	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000000
201	2	4	1	1	2	3	2	2	3	4	4	3	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	1 11
205	2	2	1	1	2	2	1	2	1	3	3	3	1	1	1	1	2	1	1	0	1	1	1	1	0	1	1	0-0	1	1	1	1	1	111111
206	2	5	1	1	2	3	3	3	3	4	3	3	2	2	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
207	2	2	1	1	2	1	2	3	3	2	3	3	1	2	2	1	2	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
208	2	3	1	2	2	3	3	3	3	4	3	3	2	2	2	3	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
211	2	4	1	1	2	3	1	2	2	2	3	3	2	2	1	1	2	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	0	101001
212	2	5	1	1	2	1	1	2	2	3	3	3	2	2	1	2	3	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	101010
213	2	4	1	2	2	2	1	2	3	3	3	3	1	1	1	2	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
215	2	2	1	2	2	3	3	3	3	4	3	3	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
216	2	2	3	1	2	1	1	2	2	1	4	3	2	1	1	3	3	1	1	1	1	1	0	1	1	1	1	1-0	1	1	1	1	1	101110
217	2	5	1	1	2	2	2	2	3	4	3	3	1	1	1	2	2	1	1	1	1	1	0	1	0	1	1	0-0	0	0	1	1	1	101011
223	2	4	2	1	2	2	2	2	2	3	4	3	2	2	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
224	2	3	3	2	2	3	3	2	2	1	3	1	1	2	1	2	3	0	0	0	1	1	0	1	1	0	0	1-0	0	0	0	0	1	101100
225	2	3	1	1	2	3	3	3	3	3	4	3	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
226	2	5	1	1	1	3	1	2	4	4	3	3	3	2	1	1	2	1	1	1	1	1	1	1	0	1	1	0-0	1	1	1	1	0	111101
229	2	4	1	2	2	3	3	3	3	3	3	2	2	3	2	1	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
230	2	5	3	2	2	2	3	4	3	1	3	3	1	1	2	2	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
233	2	3	1	2	2	1	1	2	2	2	3	2	2	2	2	3	2	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	101010
234	2	3	3	2	1	1	1	2	2	1	1	2	2	3	1	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	000010
239	2	2	1	2	2	1	3	2	3	2	1	3	3	3	1	1	4	1	1	1	1	0	1	1	1	1	1	1 1	1	1	1	1	1	111111
241	2	5	1	2	1	3	3	4	1	2	4	3	1	1	2	3	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
246	2	2	1	2	2	2	2	2	1	1	3	3	1	1	1	2	3	0	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	101111
249	2	3	1	2	2	2	2	3	3	3	3	3	3	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
253	2	3	1	2	2	2	1	3	3	4	3	3	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
254	2	4	1	2	2	2	2	3	3	4	3	1	1	1	1	2	3	0	1	0	1	1	0	1	1	1	1	1-0	1	1	0	0	1	101100
258	2	5	3	2	2	2	2	2	2	1	2	3	1	1	1	1	2	1	1	1	1	1	1	1	0	0	1	1-0	1	0	1	1	0	101001
261	2	5	1	2	2	3	3	4	2	4	3	3	2	3	1	1	4	1	1	1	0	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
266	2	5	1	2	2	2	3	4	4	3	4	2	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111101
266	2	5	3	2	2	3	3	1	2	2	3	3	2	2	1	1	2	0	1	1	1	1	0	1	1	1	1	1-0	1	1	1	1	0	101001
268	2	3	1	2	2	3	3	2	3	2	3	1	2	2	1	1	2	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	101100
269	2	3	1	1	2	1	1	2	2	3	3	1	1	1	1	2	3	0	1	0	0	1	0	1	0	0	0	1-0	0	0	0	0	1	101000
270	2	3	1	1	2	1	3	3	3	2	3	2	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
273	2	4	1	2	2	3	3	3	2	4	3	3	1	2	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
277	2	5	1	2	2	3	3	3	3	2	4	3	2	3	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
280	2	3	1	1	2	1	2	2	3	4	4	3	3	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
281	2	2	1	2	2	2	2	2	2	4	3	3	3	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
283	2	2	3	2	2	1	1	2	3	1	1	3	2	1	1	1	3	0	1	0	1	0	1	0	1	1	1	1-0	1	1	0	0	0	101110
285	2	5	1	2	2	3	3	3	2	2	1	2	1	1	2	1	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
287	2	3	1	2	2	2	2	3	2	3	3	3	3	3	2	1	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
289	2	4	1	2	2	2	3	2	3	4	4	3	2	1	1	1	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
292	2	2	1	2	2	1	2	2	2	1	2	3	1	3	2	1	2	1	1	1	1	1	1	0	1	0	1	1-0	1	0	1	1	1	111111
296	2	1	2	1	2	1	1	2	3	2	3	3	1	3	2	2	2	0	0	0	1	0	1	0	0	0	0	1-0	0	0	0	0	0	000010
297	2	2	3	2	2	3	3	2	2	2	4	3	2	1	2	1	2	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
299	2	5	3	2	2	2	3	4	2	2	3	3	2	2	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
300	2	2	1	1	2	2	2	2	2	2	3	2	1	2	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
302	2	3	1	2	1	1	2	3	2	3	1	1	2	2	2	1	3	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	000100
303	2	5	1	1	2	2	3	4	3	3	4	3	2	2	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111101
305	2	3	1	1	2	3	2	3	3	4	3	3	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111101
309	2	5	1	2	2	2	2	3	2	4	3	3	3	3	1	1	3	1	1	1	0	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
312	2	2	1	2	2	1	1	2	3	3	3	3	1	1	1	2	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
313	2	5	1	1	2	1	2	3	3	3	3	2	3	3	1	1	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
315	2	5	2	2	2	3	3	4	1	2	2	3	2	3	2	1	4	1	1	1	1	1	1	1	0	1	1	0-0	1	1	1	1	1	111111
316	2	4	1	1	2	1	2	2	4	4	4	3	1	1	2	3	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
317	2	4	1	2	2	2	3	3	3	4	2	3	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
318	2	5	1	2	2	2	1	3	2	4	3	3	3	1	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
319	2	3	1	1	2	3	1	2	4	2	3	2	2	2	1	2	2	0	1	1	0	0	0	0	0	1	0	1-0	0	0	0	0	1	101100
322	2	5	1	2	2	3	3	4	3	4	4	3	2	3	2	3	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
324	2	4	1	1	2	1	1	2	4	4	4	3	1	3	1	1	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
327	2	3	1	2	1	2	2	3	2	3	2	3	1	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
328	2	5	1	1	2	2	2	3	2	3	4	3	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
329	2	5	1	2	2	2	3	2	2	4	3	3	3	3	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
330	2	2	1	2	1	1	2	2	2	2	3	2	1	1	1	2	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	101111
331	2	5	1	2	2	2	3	4	4	4	3	2	2	1	1	2	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
333	2	3	1	1	2	2	3	2	4	4	3	3	1	1	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
336	2	4	1	1	2	2	2	3	4	4	1	1	2	3	2	1	4	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	1	101100
343	2	2	1	2	1	1	1	2	3	3	3	1	2	3	1	1	4	0	0	0	0	0	0	1	1	1	1	1-0	1	1	0	0	1	101100
347	2	3	1	2	2	2	3	1	2	2	3	3	2	3	1	2	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
351	2	5	1	1	2	1	2	3	2	4	3	3	1	1	1	3	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
353	2	3	1	1	2	1	1	2	2	2	1	3	2	3	1	1	4	0	0	1	1	0	1	1	1	1	1	1 1	1	1	0	0	1	111110
359	2	4	1	1	2	2	2	3	3	4	3	3	1	1	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
363	2	4	1	2	2	3	3	3	4	3	3	3	2	3	2	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
364	2	5	3	2	2	3	3	4	4	2	3	3	1	2	1	1	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
365	2	2	1	2	2	1	3	2	3	2	1	3	1	1	1	1	2	0	0	0	0	0	0	0	0	1	0	1-0	0	0	0	0	1	111110
366	2	3	2	2	2	1	1	2	2	1	1	3	2	3	1	1	4	0	1	0	1	0	1	1	1	1	1	1 1	1	1	0	0	1	111110
367	2	4	1	1	2	2	2	2	4	2	4	3	2	3	1	1	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	111111
368	2	2	3	1	2	1	1	2	3	1	3	2	2	3	2	1	3	0	0	0	0	0	0	0	0	0	0	1-0	0	0	0	0	0	100000
370	2	3	1	2	2	1	3	2	3	3	1	3	3	3	1	3	3	0	0	0	0	0	0	0	1	0	0	1-0	0	0	0	0	0	111111
372	2	3	1	2	2	2	2	2	2	3	3	3	2	1	2	2	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111101
375	2	4	1	1	2	3	2	2	4	3	3	3	2	2	1	2	3	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111101
381	2	5	2	2	2	2	3	2	2	1	1	2	1	1	1	1	4	1	1	0	0	1	1	1	0	1	1	1-0	1	1	0	1	1	111110
385	2	2	3	2	2	1	1	2	3	2	3	3	2	1	1	2	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
387	2	5	1	2	2	3	3	4	3	1	4	3	3	3	1	1	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
389	2	2	1	1	2	1	1	2	3	3	3	2	2	2	2	1	3	1	0	1	1	0	1	1	1	1	1	1 1	1	1	0	1	1	001110
390	2	5	1	2	2	3	3	3	3	4	3	3	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
393	2	5	1	1	2	1	3	3	4	4	3	2	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
395	2	2	1	1	2	2	2	2	4	4	3	3	1	1	2	3	4	1	1	1	1	1	1	1	1	1	1	. 1	1	1	1	1	1	111111
396	2	5	2	1	2	3	2	2	3	3	3	3	2	1	2	2	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
491	2	2	1	2	2	2	3	2	3	2	3	1	1	2	1	1	4	0	0	1	1	1	0	1	1	1	1	1-0	1	1	0	0	1	101100
408	2	5	1	1	2	3	3	4	4	4	3	3	2	1	1	2	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
410	2	5	2	1	1	1	2	2	2	1	4	3	1	1	1	1	4	0	0	0	0	0	0	1	1	0	0	1-0	0	0	0	0	1	11111
416	2	4	1	1	2	1	1	2	3	3	3	3	2	2	2	2	3	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
422	2	2	1	2	2	2	1	2	3	3	3	3	3	3	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
423	2	3	1	2	2	2	3	2	3	2	3	1	1	3	1	1	3	0	0	0	0	1	0	1	0	1	1	1-0	0	0	0	0	0	101110
425	2	5	2	2	2	3	2	2	2	1	3	3	2	2	1	1	4	1	1	0	1	1	1	1	1	1	1	1 1	1	1	1	1	1	11111
432	2	1	1	2	2	1	1	2	2	2	3	3	1	2	1	3	2	1	1	1	1	1	1	1	0	1	1	1-0	1	1	1	1	0	001000
433	2	5	2	1	2	1	3	2	4	4	3	3	2	1	1	1	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
434	2	5	1	2	2	3	3	3	4	1	3	3	1	1	1	1	4	1	1	0	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
437	2	3	3	1	2	1	2	2	3	1	3	3	1	1	1	3	4	1	1	1	1	1	1	1	1	1	1	1 1	1	1	1	1	1	111111
438	2	4	2	2	2	1	3	2	3	3	3	3	2	3	2	2	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111
442	2	2	1	2	2	2	2	3	3	3	3	3	2	3	1	2	4	1	1	1	1	1	1	1	1	1	1	0 1	1	1	1	1	1	111111

Column	Identification	Column	Identification	Column	Identification
1	Case NR	16	Purchase Type	33	2/2/2/1/1 vote w/o credit
2	Actual Response	17	Equity		
3	Age	18	Credit Rating	34	4/5 vote w/o credit
4	Marital Status	19	Judge #1 w/o credit	35	Judge #6
5	Local	20	Judge #2 w/o credit	36	Judge #7
6	Phone	21	Judge #3 w/o credit	37	Retrial Judge #1 w/credit
7	Jobtime	22	Judge #4 w/o credit		
8	Restime	23	Judge #5 w/o credit	38	Retrial Judge #2 w/credit
9	Rescar	24	Judge #1 w/credit		
10	Type Employment	25	Judge #2 w/credit	39	Retrial Judge #3 w/credit
11	Income	26	Judge #3 w/credit		
12	Bank	27	Judge #4 w/credit	40	Retrial Judge #4 w/credit
13	Financial Ref.	28	Judge #5 w/credit		
14	Loan Am..	29	Dings	41	Retrial Judge #5 w/credit
15	Loan Time	30	5/5 vote w/credit		
		31	2/2/2/1/1 vote w/credit		
		32	4/5 vote w/credit		

Exhibit 5 -- CODED PREDICTOR DATA SHEET

LOAN APPLICATION DATA FOR PURCHASE MONEY,DEALER PAPER

CASE	BANK ACCT	AGE	FIN REF	MARITAL	INCOME	LOCAL	LOAN AMT	TELEPHONE	EQUITY	JOB TIME	LOAN TERM	RES TIME	PURCHASE	RESIDENCE	CREDIT	JOB	DECISION
51060	CHECKING	31-40YRS	MAJOR	MARRIED	$6-900/MO	LOCAL	UNDER-$300	PHONE	10 OR LESS	2-5 YRS	TWO-YEARS	OVER-5YRS	NECESSITY	BUYING	POOR	SKILLED	( )
52061	CK/SAVINGS	OVER-40	MINOR	MARRIED	OVER-$900	LOCAL	$300-600	PHONE	10-20 PCT	OVER-5YRS	TWO-YEARS	UNDER-2YRS	NECESSITY	RENTING	MEDIUM	EXECUTIVE	( )
53062	NO-BANK	25-30YRS	MINOR	MARRIED	UNDER-$400	LOCAL	UNDER-$300	PHONE	10-20 PCT	2-5 YRS	TWO-YEARS	2-5 YRS	NON-NECESS	BUYING	POOR	SKILLED	( )
54063	CHECKING	22-24YRS	MAJOR	MARRIED	$4-600/MO	LOCAL	$600-900	NO-PHONE	10 OR LESS	UNDER-2YRS	THREE-YEAR	UNDER-2YRS	NON-NECESS	RENTING	UNRATED	UNSKILLED	( )
55067	CHECKING	31-40YRS	MAJOR	DIVORCED	OVER-$900	NON-LOCAL	$300-600	PHONE	10 OR LESS	UNDER-2YRS	ONE-YEAR	UNDER-2YRS	NON-NECESS	BUYING	POOR	SKILLED	( )
56069	CHECKING	31-40YRS	OTHER	MARRIED	$6-900/MO	LOCAL	UNDER-$300	PHONE	10 OR LESS	OVER-5YRS	ONE-YEAR	OVER-5YRS	NON-NECESS	RENTING	POOR	SKILLED	( )
57070	NO-BANK	21-OR-LESS	MINOR	MARRIED	$4-600/MO	LOCAL	$300-600	PHONE	10 OR LESS	UNDER-2YRS	THREE-YEAR	UNDER-2YRS	NON-NECESS	RENTING	MEDIUM	UNSKILLED	( )
58071	CHECKING	21-OR-LESS	OTHER	MARRIED	$6-900/MO	LOCAL	$300-600	PHONE	10 OR LESS	UNDER-2YRS	THREE-YEAR	OVER-5YRS	NON-NECESS	RENTING	EXCELLENT	SKILLED	( )
59072	CHECKING	31-40YRS	MINOR	MARRIED	OVER-$900	LOCAL	UNDER-$300	NO-PHONE	10 OR LESS	UNDER-2YRS	TWO-YEARS	UNDER-2YRS	NON-NECESS	BUYING	POOR	EXECUTIVE	( )
60073	CHECKING	22-24YRS	MAJOR	MARRIED	OVER-$900	LOCAL	$300-600	NO-PHONE	10 OR LESS	2-5 YRS	THREE-YEAR	2-5 YRS	NON-NECESS	BUYING	EXCELLENT	SKILLED	( )

Exhibit 6

Procedures for Lamar Data Sheets

Field Test Instruction Set and Coded Application Form

During the month of July we will be running a field-evaluation of a proposed computerized credit evaluation system. Based on extensive analysis of credit applications previously submitted to Lamar, the attached application sheet and explanation of categories have been developed.

After a regular application blank has been filled out completely, we ask that the person who would normally call in the credit report completely fill out the special credit form, being sure to put down the date, time and case number. The case number should also be put in the top left corner of the original application and an entry made in the notebook. If questions should arise relative to which category an applicant really belongs to for a given item, you should use your own best judgment based on the category definitions provided in the attached sheet.

These sheets will be picked up every few days by Mr. Gooch, who is doing the evaluation.

Case #____Date and Time of Appt.________________Salesman #_____

Age: 21 or less____ 22-24 ____ 25-30 ____ 31-40 ____ over 40 ____

Marital Status: Married ____ Divorced ____ Single ____

Local Family: No ____ Yes ____

Telephone: No ____ Yes ____

Time on Job: less than 2 years ____ 2-5 yrs. ____ over 5 years ____

Time at Residence: Less than 2 years __ 2-5 yrs. __ over 5 yrs.__

Residence Category: Rent free __ Renting __ Buying __ Own Outright__

Type Employment: Misc. __ Unskilled __ Skilled __ Executive __

Bank Acct.: None __ Savings __ Checking __ Ck/Sav.__

Finan. Ref.: Other __ Minor __ Major __

Income: Under $400__ 4-$600 __ 6-$900 __ Over $900 __

Amt. of Loan: Under $300 __ 3-$600 __ 6-$900 __ Over $900__

Equity: 10% or less __ 10-20% __ Over 20% __

Loan Term: 1 year __ 2 years __ 3 years __

Purchase type: non-necessity __ necessity __

TO BE FILLED OUT BY LOAN OFFICER

Credit Rating: Poor __ Uneval. __ Med. __ Good __

Decision: TD ___ Okay ___ Loan Officer Initials __

Date and Time of Decision ________________________________

Special Consideration? ________________________________

____

Category Explanations

1. Case #: A sequential number assigned to this application for bookkeeping purposes.
2. Date and Time of Application: Date and time application would be called in for Lamar approval.
3. Salesman #: Salesman making sale.
4. Age: Age of applicant.
5. Marital status: Current status.
6. Local Family: Close family, mother, father, children living within 50 miles of Austin. Yes, otherwise, no.
7. Telephone: Self explanatory.
8. Time on Job: Total time in current job or current job plus last previous job if previous job was the same type job.
9. Time in Residence: Total time in current home or in current home and last previous home in Austin.
10. Residence Category: Rent-free--living with parents or relatives; rest are self explanatory.
11. Type Employment:

    Misc.: Jobs which require minimum training or may be seasonal or somewhat unsteady. Examples are entertainers, retired people, people on welfare or social security, sports people, bellhops, circus and carnival workers, dishwashers, farm laborers (itinerant), garbage collectors, hod carriers, hospital aids, janitors, laundry workers, military E-3 and below, parking attendants and porters. Include pensioners, or students with an income under $250 monthly.

Unskilled: Jobs which require little training but which are steady. Examples are apprentice tradesmen, bartenders, cab drivers, clerks, clerk-stenos, collectors, cooks, dancers, factory assemblers, factory machine operators, hair dressers, laborers, longshoremen, maintenance men, merchant seaman, military enlisted E-4 and E-5, miners, moulders, musicians, oil field roustabouts, painters, paper hangers, radio announcers, salesman (under five years), service station attendants, truck drivers (short haul), typists, waiters, warehousemen and watchmen.

Skilled: Jobs normally involving considerable training skill or education. Supervision responsibilities limited. Examples are insurance adjusters, artists, auto and airplane mechanics, bookkeepers, bricklayers, bus drivers, cabinet makers, cable splicers, carpenters, chefs, civil service workers (grade WB-9 and above), credit managers, dental and laboratory technicians, die makers, draftsmen, electricians, engineers, manufacturer's representatives, firemen, food checkers, glaziers, heavy equipment operators (cranes, bulldozers, etc.), inspectors, instrument repairmen, insurance agents, interviewers, lathe operators, lithographers, machinists, mechanics, military enlisted E-7 and E-6, policemen, pressmen, salesmen (over five years), secretaries, truck drivers (long distance), barbers and carpet layers.

Executive: Normally persons with line authority supervision five or more persons. Examples are accountants, airplane pilots and navigators (commercial), architects, chemists, crew chiefs, dentists, department heads, dispatchers, editors, electronic specialists, foremen, head chefs or maitre d'hotel in large restaurant or hotel, law enforcement officers, lawyer, manager (branch, office of sales), military officers, E-8 and above, oil drillers, physicians, programmers, project engineers, railroad engineers and conductors, supervisors, superintendents and teachers.

12. Bank Account: Self explanatory.

13. Financial Reference: Previous and current creditors are listed and the following categories apply:

Code 3 (major): Credit references include major chain stores

and major banks or other large financial institutions. References do not indicate reliance on short-term, high-interest financing. Example references include: Household Finance, AVCO, GAC, Allied Finance, Sears, Wards, Scarbroughs, Joske's, Dial Finance, all banks, savings and loans and credit unions.

Code 2 (minor): Credit references include only local stores and businesses. References do not indicate reliance on short-term, high-interest financing. Examples are Term Plan, American Finance, Aetna, Republic, local department stores, jewelry stores, oil companies, etc.

Code 1 (other): Credit references include only out-of-town sources, no reference at all, or heavy dependence on short-term, high-interest financing.

14. Income: Self explanatory--use total monthly income listed for husband and wife.

15. Amount of Loan: Total amount of contract.

16. Equity: Percentage down.

17. Loan Term: (0-17 months) = 1 year, (18-27 months) = 2 years, (28-36 months) = 3 years.

18. Purchase Type: This tells whether major item in contract is necessity in a normal home or not. Examples are:

    Necessities: stove, refrigerator, washer, dryer, air conditioner, television and dishwasher.

    Non-necessities: Sewing machine, stereo.

19. Credit Rating: Judged credit rating from direct and Credit Bureau reports.

20. Decision: Approval or disapproval.

21. Loan Officer Initial: Initials of loan officer making decision.

22. Date and Time Approval or Rejection sent to: ______________

23. Special Consideration? To be annotated if special consideration such as recourse or other unusual circumstances affected decision on this application.

Exhibit 7 -- DATA FOR 150 CROSS-VALIDATION APPLICATIONS

1	310	3	1	2	1	1	2	2	2	1	1	3	2	1	2	2	1	1	1 0	2	1	1	1	1	1	1	0	53	4
2	316	5	2	2	2	2	2	3	3	3	3	1	1	2	2	2	4	2	3-0	4	3	2	2	2	2	2	1	99	51
3	311	4	1	2	2	3	3	2	3	3	3	3	1	1	1	1	4	2	3-0	1	3	2	2	2	2	2	1	50	1
4	318	4	1	2	2	3	3	3	4	3	3	4	1	3	1	2	4	2	3-0	2	2	2	2	2	2	2	1	47	1
5	310	4	1	2	2	3	3	3	3	4	3	4	2	1	2	2	4	2	3-0	1	2	2	2	2	2	2	1	48	1
6	310	4	1	1	2	1	2	2	2	3	3	1	1	2	2	2	4	2	3-0	1	3	2	2	2	2	2	1	92	48
7	310	4	1	2	2	1	1	3	3	3	3	3	2	2	3	2	1	1	3-0	2		1	1	1	1	1	0	238	196
8	310	5	1	2	1	1	1	2	2	1	2	2	1	2	2	2	2	2	3-0	2	3	1	1	1	1	1	0	90	49
9	311	2	1	2	2	1	1	2	2	2	3	1	1	3	1	2	2	2	3-0	2	3	2	2	2	2	1	1	89	47
10	311	3	1	2	2	1	2	2	2	3	3	3	1	1	2	2	3	2	3-0	1	0	2	2	2	2	2	1	46	5
11	311	5	3	2	1	1	1	3	1	3	2	1	2	1	1	2	4	2	1 1	2	0	2	2	2	2	2	1	49	49
12	313	3	1	1	2	3	1	3	2	3	3	3	3	1	3	1	3	2	3-0	4	2	2	2	2	2	2	1	50	49
13	313	2	1	2	2	1	1	2	3	3	2	2	1	2	1	2	2	1	1-0	2	2	2	2	2	2	2	1	4	4
14	315	2	1	2	2	1	1	2	3	3	3	2	4	1	3	1	4	2	1-0	1	0	2	2	2	2	2	1	45	45
15	316	4	1	2	2	1	1	3	1	3	3	1	2	1	3	2	2	1	1-0	4	0	1	2	1	2	2	0	48	48
16	510	1		1	1	3	1	2	1	3	1	1	1	1	1	2	1	1	3-0	1	0	1	1	1	1	1	0	45	1
17	510	3	1	2	2	1	1	3	2	4	3	3	1	3	1	2	1	2	5-0	4	2	1	1	1	1	1	0	3	1
18	510	5	1	2	2	3	3	4	3	1	3	3	1	1	2	2	4	2	5-0	4	2	2	2	2	2	2	1	3	1
19	512	5	1	2	2	2	1	4	1	1	3	2	1	3	1	2	4	2	3 1	1	1	2	2	2	2	2	1	3	1
20	513	1	3	2	2	1	2	1	3	3	3	4	2	3	1	2	4	2	5-0	4	2	2	2	2	2	2	1	3	1
21	514	2	1	2	2	1	1	2	2	3	1	3	1	1	1	2	1	1	1 1	2	2	1	1	1	1	1	0	1	1
22	515	3	1	2	2	1	2	2	2	1	1	1	1	1	1	2	2	2	1 1	4	0	1	1	1	1	1	0	3	1
23	6 9	2	1	1	2	1	1	2	2	1	2	2	1	1	1	1	2	1	1-0	4	3	1	1	1	1	1	0	16	1
24	6 9	3	1	1	2	3	1	2	4	3	3	4	2	1	1	2	2	2	1-0	1	1	2	2	1	2	2	1	20	5
26	615	4	1	1	2	3	1	3	3	1	3	4	1	3	1	2	1	2	5-0	2	3	1	1	1	1	1	0	1	1
27	615	4	2	2	1	3	2	2	2	1	2	1	2	1	2	2	1	1	6-0	2	3	1	1	1	1	1	0	1	1
28	710	3	1	1	2	2	1	2	3	4	3	4	1	3	1	2	4	2	5-0	4	3	2	2	2	2	2	1	23	7
29	710	2	1	2	2	1	1	2	1	3	2	2	1	3	1	2	2	2	1-0	1	2	2	2	2	2	1	1	91	77
30	710	4	1	2	2	3	3	3	3	3	3	4	2	2	2	2	3	2	5-0	2	2	2	2	2	2	2	1	18	5
31	710	2	3	2	2	1	1	1	2	3	1	1	1	1	2	1	2	2	1-0	3	0	1	1	1	1	1	0	2	1
32	712	5	1	2	2	3	3	3	3	2	3	4	4	2	1	1	3	2	5-0	1	3	2	2	2	2	2	1	4	1
34	715	5	1	1	2	1	1	3	3	3	3	3	1	3	1	2	4	2	5-0	2	3	2	2	2	2	2	1	2	1
35	715	3	1	2	2	3	1	4	3	3	1	4	2	2	2	2	4	2	5-0	1	3	2	2	2	1	1	0	2	1
36	10 9	1	3	1	2	2	1	3	3	3	2	2	1	2	2	2	2	2	1-0	2	0	1	1	2	2	1	0	65	1
37	10 9	5	1	2	2	3	3	4	2	3	3	3	2	1	3	2	4	2	1-0	1	2	2	2	2	2	2	1	54	6
38	10 9	5	1	2	2	2	3	3	2	2	3	3	1	1	2	2	4	2	5-0	5	0	2	2	2	2	2	1	50	5
39	10 9	4	1	2	2	1	1	2	4	3	3	4	2	1	3	2	4	2	5-0	3	2	2	2	2	2	2	1	60	24
40	10 9	2	1	1	2	2	1	1	2	3	3	2	2	1	2	2	4	2	5-0	2	1	2	2	2	2	2	1	50	5
41	10 9	5	1	2	2	2	3	3	3	3	3	4	2	1	2	2	4	2	5-0	4	2	2	2	2	2	2	1	56	5
42	10 10	4	1	2	2	3	1	2	2	4	3	3	3	3	2	2	1	1	5 1	5	0	1	1	1	1	1	0	76	25
43	10 10	5	1	2	2	3	2	3	3	3	3	3	2	2	2	2	4	2	1-0	2	1	2	2	2	2	2	1	45	1
44	10 10	5	1	1	2	2	2	3	3	3	3	4	2	1	1	2	4	2	4-0	0	0	2	2	2	2	2	1	46	1
45	10 10	4	1	1	2	2	1	3	4	3	3	4	1	3	1	2	4	2	1-0	0	0	2	2	2	2	2	1	44	1
46	10 11	2	1	1	1	3	1	3	3	3	3	2	1	3	1	1	4	2	5-0	2	1	2	2	2	2	2	1	47	1
47	10 11	4	1	2	2	2	1	3	3	3	3	3	1	1	2	2	4	2	1-0	2	1	2	2	2	2	2	1	60	24
48	10 11	5	1	2	2	1	2	3	3	3	3	4	2	2	1	2	4	2	6-0	2	2	2	2	2	2	2	1	1	1
49	10 12	3	1	1	2	1	1	2	4	4	3	4	1	3	1	2	4	2	5-0	2	3	2	2	2	2	2	1	2	1
50	10 12	3	1	2	2	1	1	2	2	3	3	4	2	1	2	2	4	2	5-0	4	3	2	2	2	2	2	1	2	1
51	10 13	3	1	2	2	3	2	3	3	4	3	3	2	1	1	2	4	2	6-0	2	1	2	2	2	2	2	1	2	1
52	10 15	4	3	2	2	2	3	3	1	2	3	1	1	1	1	2	1	1	6-0	3	2	1	1	1	1	1	0	2	1
53	10 15	3	1	2	2	1	1	3	3	3	3	4	1	3	1	2	4	2	5-0	1	0	2	2	2	2	2	1	18	18
54	10 15	2	1	1	2	2	1	2	3	3	3	2	2	2	2	2	4	2	5-0	1	1	2	2	2	2	2	1	18	18
55	10 16	3	1	2	2	1	2	2	1	1	1	2	2	1	3	2	1	1	4-0	1	3	1	1	1	1	1	0	16	18
56	11 9	4	1	2	2	1	1	2	2	1	1	2	3	1	2	1	2	1	0 1	2	2	1	1	1	1	1	0	43	27
57	11 9	3	1	2	2	1	1	3	2	3	3	4	1	1	2	2	4	2	5 1	4	3	2	2	2	2	2	1	19	4
58	11 10	4	1	2	2	1	1	2	3	1	3	2	3	1	2	2	4	2	1-0	2	1	2	2	2	2	2	1	1	1
59	11 10	4	2	2	2	1	1	2	1	3	3	3	4	1	2	2	1	1	5 1	2	1	1	1	1	1	1	0	4	4
60	11 12	5	1	2	2	3	3	4	1	3	3	1	1	3	1	2	4	2	5-0	2	0	2	2	2	2	2	1	4	4
62	11 12	3	1	1	2	1	1	3	4	3	3	4	2	2	2	1	4	2	5-0	1	3	2	2	2	2	2	1	23	24
63	11 13	5	1	2	2	2	1	2	2	3	2	2	1	1	1	2	1	1	1-0	4	1	1	1	1	1	1	0	2	1
64	11 14	3	1	2	2	2	1	3	3	1	3	2	2	1	3	2	4	2	5-0	2	0	2	2	2	2	2	1	19	19
65	12 10	3	1	1	2	1	1	2	2	3	3	4	2	2	2	2	4	2	5-0	4	3	2	2	2	2	2	1	24	24
66	12 14	4	1	2	2	3	2	3	3	3	3	2	2	3	3	2	2	2	1-0	1	1	2	2	2	2	2	1	50	50
67	12 15	5	3	2	1	3	3	2	3	3	2	2	1	2	2	2	4	2	6-0	5	3	2	2	2	2	2	1	19	19
68	12 16	5	1	2	2	1	3	4	2	3	2	2	2	1	2	2	4	2	6-0	0	3	2	2	2	2	2	1	21	21
69	13 9	5	1	2	1	3	3	3	1	3	3	2	2	1	2	2	1	1	1-0	2	3	1	1	1	1	1	0	17	1
70	13 9	1	3	1	2	1	1	2	2	3	2	1	1	1	1	2	2	2	3 1	0	0	1	1	1	1	1	0	40	25
71	13 12	3	1	2	2	1	1	3	3	3	2	4	2	1	2	2	4	2	4-0	2	3	2	2	2	2	2	1	1	1
73	13 13	2	3	2	1	2	2	1	2	1	1	2	2	1	2	2	2	1	4-0	1	3	1	1	1	1	1	0	1	1
74	14 9	2	3	2	2	2	2	2	3	1	2	3	2	1	2	2	2	2	2 1	4	3	1	1	2	2	1	0	23	7
75	17 10	5	1	2	2	3	3	3	4	3	3	4	2	1	1	2	4	2	1-0	1	1	2	2	2	2	2	1	51	5
76	17 10	4	2	1	2	0	1	2	0	1	2	1	1	2	1	2	1	1	5 1	4	1	1	1	1	1	1	0	51	6
77	17 10	3	3	1	2	1	2	2	3	4	3	2	1	1	1	2	4	2	1-0	2	1	2	2	2	2	2	1	50	5
78	17 10	5	1	2	2	3	2	4	3	3	3	3	1	3	1	2	4	2	1-0	4	1	2	2	2	2	2	1	50	6
79	17 10	2	1	2	2	1	1	2	2	4	3	1	2	1	3	2	4	2	1-0	4	1	2	2	2	2	2	1	49	5
80	17 10	3	1	1	2	1	1	2	3	3	3	3	1	1	1	2	4	2	1-0	0	1	2	2	2	2	2	1	49	5
81	17 10	3	1	2	2	1	1	3	4	4	3	4	2	2	1	2	4	2	1-0	2	2	2	2	2	2	2	1	49	5
82	17 10	3	1	2	2	3	2	3	3	3	3	4	1	3	1	2	4	2	1-0	2	0	2	2	2	2	2	1	45	5
83	17 10	3	1	2	2	1	1	2	3	1	3	2	1	1	2	2	1	1	1-0	4	0	1	1	1	1	1	0	40	5
84	17 10	5	1	2	2	3	2	3	3	3	3	3	1	1	2	2	4	2	1-0	0	0	2	2	2	2	2	1	5	5
85	17 10	2	1	1	2	1	1	2	2	3	1	2	1	1	2	1	2	1	1-0	0	0	1	1	1	1	1	0	45	1
86	17 10	5	1	2	2	1	3	3	3	3	2	2	1	1	2	2	3	2	1-0	2	0	2	2	2	2	2	1	5	5
87	17 11	5	1	1	2	3	1	3	4	3	3	4	2	3	1	2	4	2	1-0	4	1	2	2	2	2	2	1	4	4
88	17 13	2	1	2	2	1	1	2	3	3	3	2	1	1	3	2	4	2	1-0	4	0	2	2	2	2	2	1	24	25
90	17 14	4	1	2	2	1	1	3	3	1	3	4	1	3	1	1	4	2	1-0	1	0	2	2	2	2	2	1	1	1
91	17 14	4	1	1	2	3	1	1	3	3	3	3	1	1	2	2	4	2	1-0	2	1	2	2	2	2	2	1	20	21
92	18 9	2	1	2	2	2	1	2	2	3	3	2	1	2	1	2	4	2	1-0	2	1	2	2	2	2	2	1	17	1
93	18 9	3	1	2	2	1	1	2	2	1	3	2	1	3	1	2	2	2	1-0	4	0	1	1	1	2	2	0	21	6
94	16 11	1	1	2	1	3	1	2	2	3	3	1	1	2	2	2	2	2	6-0	1	3	1	1	2	2	2	0	4	1
95	18 11	2	1	1	1	2	1	2	2	1	2	2	2	1	2	1	1	1	1-0	2	2	1	1	1	1	1	0	23	22
96	19 9	4	1	2	2	2	2	2	2	1	2	2	2	1	3	2	1	1	1-0	4	0	1	1	1	1	1	0	23	1
97	19 9	5	3	2	1	3	2	4	2	1	2	2	1	1	2	2	1	1	1-0	2	0	1	1	1	1	1	0	22	1
98	19 9	5	1	1	2	3	2	3	3	3	3	3	1	1	1	1	1	1	1 1	2	1	1	1	1	1	1	0	16	1
99	19 9	3	1	2	2	1	3	3	3	3	3	4	2	1	2	2	4	2	1-0	0	0	2	2	2	2	2	1	16	1
100	19 9	5	1	2	2	3	2	2	3	3	3	3	2	1	2	2	4	2	1-0	4	0	2	2	2	2	2	1	1	1
101	19 14	3	1	2	2	3	1	2	1	1	3	2	1	1	2	1	2	1	1-0	1	0	1	2	1	1	2	0	3	1
102	19 14	3	1	2	2	2	2	2	2	3	2	1	1	1	2	1	1	1	1-0	2	1	1	1	1	1	1	0	4	1
103	20 9	3	3	1	2	1	1	2	2	3	1	1	1	2	1	1	2	1	1-0	2	1	2	1	1	1	1	0	47	25
104	20 10	5	3	2	2	2	2	3	3	3	3	2	1	3	1	1	3	2	1-0	4	0	2	2	2	2	2	1	24	24
105	20 10	3	1	2	2	3	1	3	3	3	3	3	1	1	2	1	4	2	5-0	1	0	2	2	2	2	2	1	2	1
106	20 13	5	1	2	2	2	1	2	3	3	3	4	2	1	2	2	4	2	5-0	2	3	2	2	2	2	2	1	1	1
107	21 10	2	1	2	1	3	2	2	2	2	3	2	2	1	3	2	4	2	5-0	0	0	2	2	2	2	2	1	22	1
108	21 10	5	2	2	2	3	2	1	2	4	3	3	2	1	2	2	4	2	5-0	1	3	2	2	2	2	2	1	2	1
110	21 10	5	1	2	2	2	2	4	3	3	2	4	1	2	2	2	1	1	5 1	0	0	1	1	1	1	1	0	95	93
111	24 9	4	1	2	1	3	2	3	2	3	3	3	2	1	3	2	3	2	3-0	1	2	2	2	2	2	2	1	1	1
112	24 11	4	1	1	2	3	1	2	4	4	3	4	2	3	1	2	2	2	5 1	2	2	2	2	2	2	2	1	30	29
113	24 11	3	1	2	2	3	1	2	4	4	3	1	1	1	1	2	4	2	4-0	2	2	2	2	2	2	2	1	30	29
114	24 11	5	1	1	2	2	3	2	2	3	3	3	2	1	2	2	3	2	4-0	4	2	2	2	2	2	2	1	2	1
115	24 11	1	1	2	2	1	2	2	2	3	1	3	1	1	2	2	4	1	5-0	4	2	2	1	2	1	1	0	25	26
116	24 11	3	2	2	2	1	3	2	3	3	3	3	2	1	3	2	2	2	6-0	3	2	1	2	1	1	2	0	3	5
117	24 12	2	1	2	2	2	2	3	2	3	3	2	2	1	3	2	4	2	5-0	2	2	2	2	2	2	2	1	1	1
118	24 13	5	1	1	2	2	3	3	3	3	3	4	2	1	2	2	4	2	5-0	0	0	2	2	2	2	2	1	20	22
119	24 13	4	1	2	2	2	1	3	2	2	3	4	2	1	1	1	3	2	5-0	4	2	2	2	2	2	2	1	1	1
121	24 17	5	1	2	2	2	3	3	2	3	3	4	2	1	2	2	4	2	6-0	0	1	2	2	2	2	2	1	22	22
122	25 9	4	1	2	2	1	1	3	2	3	3	3	2	1	2	2	4	2	5-0	0	1	2	2	2	2	2	1	6	7
123	25 9	4	1	1	2	2	2	3	3	3	3	4	1	1	1	2	4	2	4-0	0	1	2	2	2	2	2	1	6	7
124	25 9	3	1	1	2	3	1	1	3	4	3	2	2	3	1	2	3	2	2-0	0	1	2	2	2	2	2	1	6	7
125	23 14	3	2	1	2	2	1	2	2	4	3	2	1	1	2	1	3	1	5 1	4	2	2	2	2	2	2	1	1	1
127	26 9	3	1	2	2	1	1	3	4	4	3	3	1	3	1	2	4	2	5-0	4	2	2	2	2	2	2	1	2	1
128	26 9	1	1	1	2	1	1	2	2	1	2	1	1	1	2	1	2	2	1-0	2	2	1	1	1	1	1	0	3	1
129	26 9	3	1	2	2	2	2	3	3	4	3	4	1	2	2	1	4	2	5-0	2	2	2	2	2	2	2	1	1	1
130	26 13	3	1	2	1	1	1	2	2	3	3	3	1	3	1	2	4	2	6-0	4	2	2	2	2	2	2	1	2	1
132	27 9	4	2	2	2	1	1	2	3	3	3	2	2	1	3	2	1	1	1-0	1	2	1	1	1	1	1	0	5	5

13327 9	1	1	2	2	3	2	3	3	4	3	4	1	3	1	2	4	2	3-0	2	2	2	2	2	2	2	1	3	1
13427 9	4	1	1	2	3	1	3	2	3	1	4	1	1	2	2	1	1	3-0	4	4	1	1	1	1	1	0	3	1
1352712	3	1	1	2	2	2	2	3	3	3	4	1	1	1	2	4	2	5-0	1	1	2	2	2	2	2	1	25	25
1362712	4	1	2	2	3	1	4	3	3	3	4	1	3	2	1	4	2	1-0	4	1	2	2	2	2	2	1	21	21
1372713	4	1	2	1	3	3	2	2	1	2	2	2	2	2	2	2	2	4-0	0	0	1	2	2	2	1	0	22	21
1382713	5	1	2	2	1	2	3	1	3	3	3	2	1	3	2	1	1	8-0	1	2	1	1	1	1	1	0	1	1
1392715	1	3	2	2	2	3	1	2	4	3	2	1	1	2	1	4	2	1-0	4	8	2	2	2	2	2	1	18	18
14028 9	2	3	2	2	2	3	2	3	4	2	2	2	1	2	2	3	2	1-0	3	1	2	2	2	2	1	1	3	1
14129 9	4	1	2	2	2	1	2	2	3	2	2	1	1	2	1	1	1	1-0	1	1	1	1	1	1	1	0	2	1
1422810	3	1	1	2	1	1	2	3	4	3	4	2	3	3	2	2	2	1-0	4	1	1	2	1	2	1	0	7	6
1432811	4	1	1	2	1	2	3	3	3	3	4	1	1	2	2	3	2	1-0	4	0	2	2	2	2	2	1	3	1
1452816	1	1	1	2	1	1	2	2	3	2	2	1	1	2	2	1	1	1-0	1	2	1	1	1	1	1	0	3	1
1462817	5	1	2	2	3	2	2	3	2	3	4	3	1	3	1	1	1	1-0	1	2	1	1	1	1	1	0	65	65
14831 9	3	1	1	2	1	[illegible]	2	[illegible]	[illegible]	3	4	1	2	2	2	4	2	5-0	4	1	2	2	2	2	2	1	50	1
1493110	4	1	1	2	1	2	3	[illegible]	3	3	3	2	1	2	2	4	2	1-0	2	2	2	2	2	2	2	1	53	7
1503110	5	1	1	1	2	2	4	3	3	3	3	1	3	1	2	4	2	6-0	3	2	2	2	2	2	2	1	50	5
15131 9	5	1	2	2	3	1	2	3	4	3	4	2	2	1	2	3	2	6-0	2	2	2	2	2	2	2	1	46	1
15231 7	5	1	2	2	3	1	4	1	3	3	2	2	1	1	2	4	2	3-0	2	2	2	2	2	2	2	1	46	6
15331 9	5	3	2	2	3	3	4	3	4	3	3	4	3	1	2	4	2	5-0	2	1	2	2	2	2	2	1	44	1
1543110	2	1	2	2	2	1	2	3	4	3	3	1	1	1	2	4	2	1-0	1	2	2	2	2	2	2	1	49	6
1573111	4	1	1	2	2	3	2	3	3	3	3	1	1	1	2	4	2	1-0	2	0	2	2	2	2	2	1	7	5
1593111	5	1	1	2	3	3	3	4	3	3	4	2	1	1	2	3	2	1-0	1	0	2	2	2	2	2	1	1	1
1623114	5	3	2	2	3	3	3	3	3	3	3	3	3	2	2	4	2	5-0	0	0	2	2	2	2	2	1	20	28
1633114	3	1	2	2	2	1	2	3	3	3	2	2	1	2	2	4	2	6-0	0	0	2	2	2	2	2	1	25	26
1643115	4	1	2	2	3	2	2	3	4	3	4	3	1	2	2	2	2	2 1	3	0	2	2	2	2	2	1	2	1
1653115	5	2	2	2	2	3	3	1	3	3	1	1	1	3	2	3	1	1-0	1	1	1	2	2	2	2	1	18	19
1663116	5	1	2	2	1	1	1	1	4	2	1	2	2	1	2	2	2	5 1	3	0	1	2	1	1	1	0	18	18

Column	Identification
1	Case NR.
2	Day Sold
3-4	Hr. Sold
5	Age
6	Marital
7	Local
8	Telephone
9	Time on Job
10	Time in Residence
11	Residence Category
12	Type Employment
13	Bank Account
14	Financial Ref.
15	Income
16	Loan Amount
17	Loan Term
18	Purchase Type
19	Credit Rating
20	Decision
21	Loan Officer Making Decision
22	Special Considerations
23	Salesman
24	Clerk Coding Data
25	Judge #1 w/credit
26	Judge #2 w/credit
27	Judge #3 w/credit
28	Judge #4 w/credit
29	Judge #5 w/credit
30	4/5 vote w/credit
31-32-33	Hrs. in System
34-35-36	Hrs. in Credit System

Exhibit 8 -- AID4UT OUTPUT FOR JUDGE #1

```
GROUP    N     TOTAL WEIGHT        SUM OF Y        SUM Y-SQUARE          T S S            MEAN           STD. DEV.
  1     224    224.00000         99.000000         99.000000         55.245536        .44196429        .49662043


                                                   SPLIT SUMMARY

SPLIT GROUP    1 ON PREDICTOR  16    CRERAT                INTO GROUP    2 WITH CODES  0
                                                           AND GROUP     3 WITH CODES  1  2  3
      GROUP    2  N=     76 MEAN=     .01        GROUP   3  N=   148 MEAN=      .66     T=  11.74      RSQ=  .38285 F-RSQ=  137.7
*****************************************************************************************************************************
SPLIT GROUP    3 ON PREDICTOR  11    FINREF                INTO GROUP    4 WITH CODES  0
                                                           AND GROUP     5 WITH CODES  1  2
      GROUP    4  N=     19 MEAN=    0.00        GROUP   5  N=   129 MEAN=      .76     T=  7.698      RSQ=  .55585 F-RSQ=  86.08
*****************************************************************************************************************************
SPLIT GROUP    5 ON PREDICTOR  10    BANKAC                INTO GROUP    6 WITH CODES  0
                                                           AND GROUP     7 WITH CODES  1  3  2
      GROUP    6  N=     30 MEAN=     .37        GROUP   7  N=    99 MEAN=      .88     T=  6.617      RSQ=  .66515 F-RSQ=  71.81
*****************************************************************************************************************************
SPLIT GROUP    7 ON PREDICTOR  16    CRERAT                INTO GROUP    8 WITH CODES  2  1
                                                           AND GROUP     9 WITH CODES  3
      GROUP    8  N=     46 MEAN=     .76        GROUP   9  N=    53 MEAN=      .98     T=  3.520      RSQ=  .68678 F-RSQ=  15.12
*****************************************************************************************************************************
SPLIT GROUP    8 ON PREDICTOR  12    LOANAH                INTO GROUP   10 WITH CODES  0
                                                           AND GROUP    11 WITH CODES  1  2
      GROUP   10  N=     16 MEAN=     .94        GROUP  11  N=    30 MEAN=      .67     T=  2.104      RSQ=  .70063 F-RSQ=  10.09
*****************************************************************************************************************************
SPLIT GROUP    6 ON PREDICTOR  16    CRERAT                INTO GROUP   12 WITH CODES  1
                                                           AND GROUP    13 WITH CODES  2  3
      GROUP   12  N=     19 MEAN=     .16        GROUP  13  N=    11 MEAN=      .73     T=  3.665      RSQ=  .74151 F-RSQ=  34.32
*****************************************************************************************************************************
SPLIT GROUP   11 ON PREDICTOR   2    MARITL                INTO GROUP   14 WITH CODES  2
                                                           AND GROUP    15 WITH CODES  0  1
      GROUP   14  N=      5 MEAN=     .20        GROUP  15  N=    25 MEAN=      .76     T=  2.613      RSQ=  .76516 F-RSQ=  21.75
*****************************************************************************************************************************


*****************************************************************************************************************************

                                                   FINAL SUMMARY


NCF     TOTAL TSS        TOTAL BSS        TOTAL WSS        R-SQUARED        R                        F-ANOVA    DF1    DF2

 8      55.245536        42.271928        12.973608        .76516459     .87474                      100.5419     7     216

*****************************************************************************************************************************
```

ANALYSIS WITH . BSS/TSS(I)

TRIAL	1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	.012652	.005742	.004474	.050844	.067264	.012130
2	.031450	.038275	.000476	.116372	.048102	.019809
3	.040572	.052305	.000871	.161468	.111939	.062779
4	.005522	.068966	.048563	.081904	.056162	.100107
5	.005210	.078742	.064935	.056104	.010019	.078099
6	.002051	.023923	0.000000	.210526	.011164	.003190
7	.077640	.196903	.060000	.064000	.083333	.110869
8	.027215	0.000000	.044085	0.000000	.028864	.095751
9	.014815	.066530	0.000000	.109375	.066964	.005919
10	.022222	0.000000	0.000000	0.000000	0.000000	0.000000
11	.015619	.002256	.011302	.003011	.017374	.030725
12	.027090	.126374	.027098	0.000000	.108173	.018519
13	.146667	0.000000	.066667	0.000000	.085714	.030303
SUBSUM	.43774	.67912	.34775	.86368	.69307	.56828
TRIALS	13	13	13	13	13	13
MEANBS	.03367	.05224	.02675	.06644	.05331	.04371
S D	.03804	.05695	.02902	.06588	.03648	.03871
CF VAR	1.12979	1.09019	1.08469	.99163	.68422	.88565
VAR	.00145	.00324	.00084	.00434	.00133	.00150
RECSUM	.43774	.67912	.34775	.86368	.69307	.56828

ANALYSIS WITH BSS/TSS(I)

TRIAL	7RESCAT	8TYPEMP	9INCOME	10BANKAC	11FINREF	12LOANAM
1	.055071	.043629	.032089	.124566	.171450	.022689
2	.111324	.038343	.046130	.198367	.288682	.015106
3	.133523	.066338	.060806	.256398	.122828	.021732
4	.076468	.042971	.068569	.304772	.030917	.019025
5	.067625	.029553	.033586	.011864	.039589	.091450
6	0.000000	.027838	.091453	0.000000	.019253	.039075
7	.095238	.109375	.142857	.012422	.045455	.100308
8	.071637	.093567	.078947	0.000000	.075815	.078947
9	0.000000	.157443	.066964	0.000000	.066964	.014815
10	0.000000	0.500000	0.900000	0.000000	.060856	0.000000
11	.007770	.030725	.014815	.014854	.013333	.009697
12	.029304	.053571	.184615	.053571	.002455	.040724
13	.030303	.146667	.111111	0.000000	0.000000	0.000000
SUBSUM	.67926	.84002	.93274	.67601	.94302	.45326
TRIALS	13	13	13	10	13	12
MEANBS	.05225	.06462	.07175	.06760	.07260	.03777
S D	.04327	.04526	.04957	.05848	.07807	.03234
CF VAR	.82813	.71592	.69095	1.30889	1.07539	.85623
VAR	.00187	.00214	.00246	.00383	.00610	.00105
RECSUM	.67926	.84002	.93274	.87882	.94382	.49103

ANALYSIS WITH BSS/TSS(I)

	13LOANTM	14NECESS	15EQUITY	16CRERAT
1	.004928	.037848	.001697	.382847
2	.009893	.029246	.000391	.187277
3	.001791	.009781	.000053	.215195
4	.003382	.007697	.000475	.113296
5	.010186	.012650	.000346	.000047
6	.014440	.000109	.043062	.324192
7	.010000	.036199	0.000000	0.000000
8	.026864	.014582	.000258	.009311
9	.066966	.106838	0.000000	0.000000
10	.022222	0.000000	0.000000	0.000000
11	.027280	.003011	.004138	0.000000
12	.031731	.010747	.008316	0.000000
13	.051852	0.000000	.085714	.066667
SUBSUM	.28144	.26783	.14445	1.24883
TRIALS	13	13	13	10
MEANBS	.02165	.02060	.01111	.12988
S D	.01879	.02785	.02432	.13481
CF VAR	.86801	1.35195	2.18827	1.03790
VAR	.00035	.00078	.00059	.01817
RECSUM	.28144	.26783	.14445	1.68848

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 13 GRAND MEAN = .671096 ROUND CRIT. = .010TSST = 55.2

		SUB(VARIANT) SUMBSS/TSS(I)	RECONSTRUCTED SUMBSS/TSS(I)	S D	COEFF VAR	VARIANCE PCT	GRJ MEAN	N GROUPS	RANK
1	AGE	.43774	.43774	.03804	1.12979	.00145	65.22756	13	12.0
2	MARITL	.67912	.67912	.05695	1.09019	.00324	10..19599	13	9.0
3	LOCAL	.34775	.34775	.02902	1.08469	.00084	51.61832	13	13.0
4	PHONE	.86368	.86368	.06588	.94163	.00434	1.8.69756	13	5.0
5	JOBTIM	.69307	.69307	.03648	.68422	.00133	133.27484	13	7.0
6	RESTIM	.56828	.56828	.03.71	.88565	.00150	84.67912	13	10.0
7	RESCAT	.67926	.67926	.04327	.82813	.00187	101.21789	13	8.0
8	TYPEMP	.84002	.84002	.04628	.71592	.00214	125.17098	13	6.0
9	INCOME	.93274	.93274	.04957	.69095	.00246	138.98699	13	3.0
10	BANKAC	.67601	.87882	.08848	1.30889	.00783	130.95290	10	4.0
11	FINREF	.94382	.94382	.07807	1.07539	.00610	140.63803	13	2.0
12	LOANAM	.45326	.49103	.03234	.85623	.00105	73.16860	12	11.0
13	LOANTM	.28144	.28144	.01879	.86801	.00035	41.93795	13	14.0
14	NECESS	.26783	.26783	.02785	1.35195	.00078	39.90897	13	15.0
15	EQUITY	.14445	.14445	.02432	2.18827	.00059	21.52460	13	16.0
16	CRERAT	1.29883	1.68848	.13481	1.03790	.01817	251.60051	10	1.0

```
GROUP   3        MEAN=      .66 RSQ =  .383
N=   148         S.D.=      .47 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  1  2  3
```

```
GROUP   5        MEAN=      .76 RSQ =  .556
N=   129         S.D.=      .43 PROB= 0.000
PREDICTOR 11 FINREF
CODES  1  2
```

```
GROUP   7        MEAN=      .88 RSQ =  .665
N=    99         S.D.=      .33 PROB= 0.000
PREDICTOR 10 BANKAC
CODES  1  3  2
```

```
GROUP   6        MEAN=      .37 RSQ =  .665
N=    30         S.D.=      .48 PROB= 0.000
PREDICTOR 10 BANKAC
CODES  0
```

```
GROUP   4=FINAL MEAN=      6.06 RSQ =  .556
N=    19         S.D.=     0.00 PROB= 0.000
PREDICTOR 11 FINREF
CODES  0
```

```
GROUP   1        MEAN=      .44
N=   224         S.D.=      .50
        LEVEL 1
```

```
GROUP   2=FINAL MEAN=      .01 RSQ =  .383
N=    76         S.D.=      .11 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  0
```

```
GROUP  13=FINAL MEAN=      .73 RSQ =  .742
N=    11         S.D.=      .45 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  2  3
```

```
GROUP   6        MEAN=      .37
N=    30         S.D.=      .48
        LEVEL 2
```

```
GROUP  12=FINAL MEAN=      .16 RSQ =  .742
N=    19         S.D.=      .36 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  1
```

```
GROUP    9*FINAL MEAN=      .98 RSQ =  .687
 N=     53        S.D.=      .14 PROB=  .002
 PREDICTOR 16 CRERAT
 CODES  3

GROUP    7        MEAN=      .88
 N=     99        S.D.=      .33
         LEVEL 2

GROUP   10*FINAL MEAN=      .94 RSQ =  .701
 N=     16        S.D.=      .24 PROB=  .002
 PREDICTOR 12 LOANAM
 CODES  0

GROUP    8        MEAN=      .76 RSQ =  .687
 N=     46        S.D.=      .43 PROB=  .000
 PREDICTOR 16 CRERAT
 CODES  2  1

GROUP   11        MEAN=      .67 RSQ =  .701
 N=     30        S.D.=      .47 PROB=  .002
 PREDICTOR 12 LOANAM
 CODES  1  2

GROUP   15*FINAL MEAN=      .76 RSQ =  .765
 N=     25        S.D.=      .43 PROB= 0.000
 PREDICTOR  2 MARITL
 CODES  0  1

GROUP   14*FINAL MEAN=      .20 RSQ =  .765
 N=      5        S.D.=      .40 PROB= 0.000
 PREDICTOR  2 MARITL
 CODES  2
```

Exhibit 9 -- AID4UT OUTPUT FOR JUDGE #2

```
GROUP    N   TOTAL WEIGHT     SUM OF Y     SUM Y-SQUARE       T S S         MEAN        STD. DEV.
  1    224   224.00000       107.00000     107.00000       55.888393     .47767857    .49950151


                                         SPLIT SUMMARY

SPLIT GROUP   1 ON PREDICTOR  16   CRERAT            INTO GROUP    2 WITH CODES  0  1
                                                     AND GROUP     3 WITH CODES  2  3
      GROUP   2  N=   116 MEAN=     .13       GROUP   3  N=   108 MEAN=     .85    T=  15.58     RSQ=  .52264 F-RSQ=  242.9
****************************************************************************************************************************
SPLIT GROUP   3 ON PREDICTOR  10   BANKAC            INTO GROUP    4 WITH CODES  0
                                                     AND GROUP     5 WITH CODES  1  2  3
      GROUP   4  N=    17 MEAN=     .53       GROUP   5  N=    91 MEAN=     .91    T=  4.391     RSQ=  .55997 F-RSQ=  18.85
****************************************************************************************************************************
SPLIT GROUP   2 ON PREDICTOR  14   CRERAT            INTO GROUP    6 WITH CODES  0
                                                     AND GROUP     7 WITH CODES  1
      GROUP   6  N=    76 MEAN=    0.00       GROUP   7  N=    40 MEAN=     .38    T=  8.694     RSQ=  .62592 F-RSQ=  38.78
****************************************************************************************************************************
SPLIT GROUP   7 ON PREDICTOR  13   BANKAC            INTO GROUP    8 WITH CODES  1  0
                                                     AND GROUP     9 WITH CODES  3  2
      GROUP   8  N=    21 MEAN=     .10       GROUP   9  N=    19 MEAN=     .68    T=  4.715     RSQ=  .68783 F-RSQ=  43.43
****************************************************************************************************************************
SPLIT GROUP   5 ON PREDICTOR  11   FINREF            INTO GROUP   10 WITH CODES  1  0
                                                     AND GROUP    11 WITH CODES  2
      GROUP  10  N=    23 MEAN=     .74       GROUP  11  N=    68 MEAN=     .97    T=  3.585     RSQ=  .70430 F-RSQ=  12.15
****************************************************************************************************************************
SPLIT GROUP  10 ON PREDICTOR   5   JOBTIM            INTO GROUP   12 WITH CODES  0  1
                                                     AND GROUP    13 WITH CODES  2
      GROUP  12  N=    17 MEAN=     .88       GROUP  13  N=     6 MEAN=     .33    T=  3.010     RSQ=  .72822 F-RSQ=  19.16
****************************************************************************************************************************
SPLIT GROUP   4 ON PREDICTOR   6   INCOME            INTO GROUP   14 WITH CODES  0  1
                                                     AND GROUP    15 WITH CODES  2  3
      GROUP  14  N=    11 MEAN=     .73       GROUP  15  N=     6 MEAN=     .17    T=  2.464     RSQ=  .75005 F-RSQ=  18.87
****************************************************************************************************************************
SPLIT GROUP   9 ON PREDICTOR   1   AGE               INTO GROUP   16 WITH CODES  0  1  2  3
                                                     AND GROUP    17 WITH CODES  4
      GROUP  16  N=    14 MEAN=     .57       GROUP  17  N=     5 MEAN=    1.00    T=  1.032     RSQ=  .76216 F-RSQ=  10.95
****************************************************************************************************************************
SPLIT GROUP  14 ON PREDICTOR  16   CRERAT            INTO GROUP   18 WITH CODES  2
                                                     AND GROUP    19 WITH CODES  3
      GROUP  18  N=     6 MEAN=     .50       GROUP  19  N=     5 MEAN=    1.00    T=  2.023     RSQ=  .77436 F-RSQ=  11.57
****************************************************************************************************************************


****************************************************************************************************************************

                                         FINAL SUMMARY

NCF     TOTAL TSS      TOTAL BSS      TOTAL WSS      R-SQUARED       R                       F-ANOVA    DF1    DF2

 10     55.888393      43.277749      12.610644      .77436023    .87998                     81.6016     9     214

****************************************************************************************************************************
```

ANALYSI [illegible] BSS/TSS(I)

TRIAL	1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	.021757	.001126	.000001	.050085	.102986	.030895
2	.012544	.017746	.001126	.084359	.077573	.041 43
3	.010103	.020321	.007489	.005933	.028640	.001160
4	.164835	.018657	.088645	.060397	.058398	.032997
5	.016118	.008032	.003485	.031581	.024549	.084388
6	.036529	0.000000	.014781	0.000000	.301423	.063813
7	.041246	.122454	.001894	0.000000	.173611	0.000000
8	.164835	0.000000	.011801	0.000000	.072485	.068946
9	.083333	0.000000	.015625	0.000000	.015625	.027778
10	0.000000	0.000000	0.000000	0.000000	.312500	0.000000
11	.018690	.007163	.046716	.063864	.107071	.116883
12	.263158	.039803	0.000000	.052632	.042105	.263158
13	.244444	0.000000	.000463	0.000000	.093333	.027222
SUSSUM	[illegible]	.24481	.19132	.34885	1.44029	.75873
TRIALS	[illegible]	13	13	13	13	13
MEANBS	[illegible]	.01883	.01472	.02683	.11079	.05836
S D	[illegible]	.03239	.02455	.03012	.07298	.06786
CF VAR	1.04559	1.71999	1.66808	1.12238	[illegible]	1.16262
VAR	.00811	.00105	.00060	.00091	[illegible]	.00460
RECSUM	1.07717	.24481	.19132	.34885	1.44029	.75873

ANALYSIS WITH BSS/TSS(I)

TRIAL	7RESCAT	8TYPEMP	9INCOME	10BANKAC	11FINREF	12LOANAM
1	.057134	.023335	.042201	.177256	.118473	.010966
2	.079855	.009034	.026719	.153902	.152692	.016644
3	.080720	.015103	.014768	.069184	.053465	.008720
4	.064052	.028913	.018315	.369089	.216000	.016835
5	.098527	.031589	.021717	.004096	.126186	.047403
6	.188235	.036029	.017507	0.000000	.004980	.138235
7	0.000000	.181481	.288089	0.000000	.084175	.001694
8	.022161	.022161	0.000000	0.000000	.133516	.089739
9	0.000000	.066667	0.000000	0.000000	.066667	.118519
10	0.000000	0.000000	.022222	0.000000	0.000000	0.000000
11	.038384	.067821	.019956	.011750	0.000000	.014433
12	0.000000	.078947	.095694	0.000000	.023684	.095694
13	.093333	.093333	.012626	0.000000	.004286	.093333
SUBSUM	.64244	.65442	.57981	.76528	.98412	.68547
TRIALS	13	13	13	14	12	13
MEANBS	.04942	.05034	.04460	.07139	.08201	.05273
S D	.05394	.04679	.07399	.11290	.06585	.05607
CF VAR	1.09161	.92943	1.65895	1.58150	.80299	1.06329
VAR	.00291	.00219	.00547	.01275	.00434	.00314
RECSUM	.64244	.65442	.57981	.92806	1.06614	.68547

ANALYSIS WITH BSS/TSS(I)

	13LOANTM	14NECESS	15EQUITY	16CRERAT
1	.087768	.257678	.006811	.522442
2	.031464	.002059	.011423	.143439
3	.036753	.002298	.006449	.282178
4	.182526	.000209	.008000	0.000000
5	.047483	.000506	.000290	.097929
6	.029789	.063813	.016244	.074774
7	.032693	0.000000	0.000000	.202231
8	0.000000	0.000000	.011722	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	.022222	0.000000	0.000000	.312500
11	.018759	.016529	.004040	.067821
12	.078947	.013158	.037895	0.000000
13	.118519	0.000000	0.000000	.000463
SUBSUM	.60666	.10624	.10607	1.70378
TRIALS	13	13	13	9
MEANBS	.04666	.00817	.00816	.18931
S D	.05017	.01691	.00874	.15237
CF VAR	1.07518	2.06957	1.07093	.80486
VAR	.00252	.00029	.00008	.02322
RECSUM	.60666	.10624	.10607	2.46101

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 13 GRAND MEAN = .743593 ROUND CRIT. = .6107887 = 55.9

		SUB(VARIANT) SUMBSS/TSS(I)	RECONSTRUCTED SUMBSS/TSS(I)	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	AGE	1.07717	1.07717	.09083	1.08659	.00811	144.06072	13	3.0
2	MARITL	.24481	.24481	.03239	1.71999	.03105	32.92278	13	13.0
3	LOCAL	.19132	.19132	.02455	1.66008	.00060	25.72964	13	14.0
4	PHONE	.34885	.34885	.03012	1.12238	.00091	46.91624	13	12.0
5	JOBTIM	1.44029	1.44029	.09298	.83923	.00865	193.69334	13	2.0
6	RESTIM	.75873	.75873	.06786	1.10263	.00460	102.03581	13	6.0
7	RESCAT	.64244	.64244	.05394	1.09141	.00291	86.39702	13	9.0
8	TYPEMP	.65442	.65442	.04679	.92943	.00219	88.00723	13	8.0
9	INCOME	.57981	.57981	.07399	1.65895	.00547	77.77440	13	11.0
10	BANKAC	.78528	.92026	.11290	1.58150	.01275	124.00699	11	5.0
11	FINREF	.98412	1.06614	.06585	.80299	.00434	143.37629	12	4.0
12	LOANAM	.68547	.68547	.05607	1.06329	.00314	92.18411	13	7.0
13	LOANTM	.60666	.60666	.05017	1.07518	.00252	81.58251	13	10.0
14	NECESS	.10624	.10624	.01691	2.06957	.00029	14.28762	13	15.0
15	EQUITY	.10607	.10607	.02874	1.07093	.00080	14.26504	13	16.0
16	CRERAT	1.70378	2.46101	.15237	.80486	.02322	330.96227	9	1.0

```
GROUP 1        MEAN=    .46
N=  224        S.D.=    .50
     LEVEL 1

GROUP 3        MEAN=    .85 RSQ = .522
N=  108        S.D.=    .36 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 2 3

GROUP 2        MEAN=    .13 RSQ = .522
N=  116        S.D.=    .34 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 0 1

GROUP 5        MEAN=    .91 RSQ = .560
N=   91        S.D.=    .28 PROB= .000
PREDICTOR 10 BANKAC
CODES 1 2 3

GROUP 4        MEAN=    .53 RSQ = .560
N=   17        S.D.=    .50 PROB= .000
PREDICTOR 10 BANKAC
CODES 0

GROUP 7        MEAN=    .36 RSQ = .626
N=   40        S.D.=    .48 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 1

GROUP 6*FINAL MEAN=     0.09 RSQ = .626
N=   76        S.D.=    0.00 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 0

GROUP 11*FINAL MEAN=    .97 RSQ = .784
N=   68        S.D.=    .17 PROB= .061
PREDICTOR 11 FINREF
CODES 2

GROUP 10       MEAN=    .74 RSQ = .784
N=   23        S.D.=    .44 PROB= .001
PREDICTOR 11 FINREF
CODES 1 0

GROUP 14       MEAN=    .73 RSQ = .750
N=   11        S.D.=    .45 PROB= .000
PREDICTOR 9 INCOME
CODES 0 1

GROUP 15*FINAL MEAN=    .17 RSQ = .750
N=    6        S.D.=    .37 PROB= .000
PREDICTOR 9 INCOME
CODES 2 3

GROUP 9        MEAN=    .48 RSQ = .688
N=   19        S.D.=    .46 PROB= 0.000
PREDICTOR 10 BANKAC
CODES 3 2

GROUP 8*FINAL MEAN=     .10 RSQ = .688
N=   21        S.D.=    .29 PROB= 0.000
PREDICTOR 10 BANKAC
CODES 1 0
```

```
GROUP 19*FINAL MEAN=    1.00 RSQ =  .774
N=    5        S.D.=    0.00 PROB=  .001
PREDICTOR 16 CRERAT
CODES  3
```

```
GROUP 14       MEAN=     .73
N=   11        S.D.=     .45
        LEVEL 2
```

```
GROUP 18*FINAL MEAN=     .50 RSQ =  .774
N=    6        S.D.=     .50 PROB=  .001
PREDICTOR 16 CRERAT
CODES  2
```

```
GROUP 12*FINAL MEAN=     .88 RSQ =  .728
N=   17        S.D.=     .32 PROB=  .000
PREDICTOR  5 JOBTIN
CODES  0  1
```

```
GROUP 10       MEAN=     .74
N=   23        S.D.=     .44
        LEVEL 2
```

```
GROUP 13*FINAL MEAN=     .33 RSQ =  .728
N=    6        S.D.=     .47 PROB=  .000
PREDICTOR  5 JOBTIN
CODES  2
```

```
GROUP 17*FINAL MEAN=    1.00 RSQ =  .762
N=    5        S.D.=    0.00 PROB=  .001
PREDICTOR  1 AGE
CODES  4
```

```
GROUP  9       MEAN=     .68
N=   19        S.D.=     .46
        LEVEL 2
```

```
GROUP 16*FINAL MEAN=     .57 RSQ =  .762
N=   14        S.D.=     .49 PROB=  .001
PREDICTOR  1 AGE
CODES  0  1  2  3
```

Exhibit 10 -- AID4UT OUTPUT FOR JUDGE #3

```
GROUP   N    TOTAL WEIGHT    SUM OF Y     SUM Y-SQUARE    T S S       MEAN        STD. DEV.
  1    224   224.00000       104.00000    104.00000       55.714286   .46428571   .49872286


                                        SPLIT SUMMARY

SPLIT GROUP   1 ON PREDICTOR  16   CRERAT          INTO GROUP   2 WITH CODES  0  1
                                                   AND GROUP    3 WITH CODES  2  3
      GROUP   2  N=   116 MEAN=    .09      GROUP   3  N=  108 MEAN=    .86   T=  17.85   RSQ=  .58945 F-RSQ=  318.7
**************************************************************************************************************
SPLIT GROUP   3 ON PREDICTOR  11   FINREF          INTO GROUP   4 WITH CODES  0
                                                   AND GROUP    5 WITH CODES  1  2
      GROUP   4  N=    15 MEAN=    .47      GROUP   5  N=   93 MEAN=    .92   T=  5.305   RSQ=  .63809 F-RSQ=  29.71
**************************************************************************************************************
SPLIT GROUP   2 ON PREDICTOR  16   CRERAT          INTO GROUP   6 WITH CODES  0
                                                   AND GROUP    7 WITH CODES  1
      GROUP   6  N=    76 MEAN=    .01      GROUP   7  N=   40 MEAN=    .25   T=  4.444   RSQ=  .66648 F-RSQ=  17.39
**************************************************************************************************************
SPLIT GROUP   7 ON PREDICTOR  10   BANKAC          INTO GROUP   8 WITH CODES  0
                                                   AND GROUP    9 WITH CODES  2  1  3
      GROUP   8  N=    20 MEAN=    .05      GROUP   9  N=   20 MEAN=    .45   T=  3.210   RSQ=  .69320 F-RSQ=  20.50
**************************************************************************************************************
SPLIT GROUP   5 ON PREDICTOR   4   PHONE           INTO GROUP  10 WITH CODES  0
                                                   AND GROUP   11 WITH CODES  1
      GROUP  10  N=     8 MEAN=    .63      GROUP  11  N=   85 MEAN=    .95   T=  3.547   RSQ=  .70731 F-RSQ=  10.51
**************************************************************************************************************
SPLIT GROUP   9 ON PREDICTOR   3   TYPEMP          INTO GROUP  12 WITH CODES  0  2  3
                                                   AND GROUP   13 WITH CODES  1
      GROUP  12  N=     9 MEAN=    .22      GROUP  13  N=   11 MEAN=    .64   T=  1.936   RSQ=  .72255 F-RSQ=  11.92
**************************************************************************************************************


**************************************************************************************************************
                                        FINAL SUMMARY

NCF   TOTAL TSS    TOTAL BSS    TOTAL WSS    R-SQUARED    R         F-ANOVA   DF1   DF2
 7    55.714286    40.256335    15.457950    .72254961    .85003    94.1870    6    217
**************************************************************************************************************
```

ANALYSIS WITH BSS/TSS(I)

TRIAL	1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	.037169	.002888	.000169	.030352	.008457	.014256
2	.003226	.022145	.063168	.094477	.035704	.001122
3	.013178	.022819	.006505	.092028	.011723	.006262
4	.028833	.010753	.032967	.009406	.040000	.030303
5	.016679	.033262	.010237	.121480	.025562	.014366
6	.015152	0.000000	.000.02	0.000000	.032079	.044995
7	.034091	.111772	.002058	0.000000	.010082	.014973
8	.002976	0.000000	.002976	0.000000	.142857	.008929
9	.004762	0.000000	0.000000	0.000000	.004762	.004762
10	[illegible]	.002256	.015619	.059048	.010233	.028889
11	.122807	.028340	0.000000	.097744	.122807	.043062
SUBSUM	.28709	.23422	.07379	.41953	.50427	.21192
TRIALS	11	11	11	10	11	11
MEANS	.02610	.02129	.00671	.04195	.04584	.01927
S D	.03317	.03098	.00956	.04524	.04470	.01454
CF VAR	1.27081	1.45477	1.42561	1.07831	.97517	.75476
VAR	.00110	.00096	.00009	.00205	.00200	.00021
RECSUM	.28705	.23422	.07379	.46149	.50427	.21192

ANALYSIS WITH BSS/TSS(I)

TRIAL	7RESCAT	8TYPEMP	9INCOME	10BANKAC	11FINREF	12LOANAM
1	.041071	.032521	.031554	.132947	.115949	.043061
2	.044859	.052367	.043378	.119250	.209823	.005667
3	.003223	.035904	.047188	.050266	.042424	.012290
4	.006536	.084175	.096774	.213333	.120000	.030303
5	.049118	.020167	.049205	.101214	.067406	.012959
6	.030303	.171513	0.000000	0.000006	.084175	.015152
7	.031339	.039545	.052441	.052441	.027778	.000110
8	0.000000	.036584	.002976	.047619	0.000080	.005102
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	.007778	.012650	.012000	.155550	.013333	.009697
11	.027544	.043062	.052632	0.000000	.022556	.052632
SUBSUM	.23177	.59049	.38815	.87963	.70352	.19400
TRIALS	11	10	11	10	10	11
MEANS	.02107	.05405	.03529	.08796	.07035	.01764
S D	.01796	.04268	.02839	.06531	.06105	.01613
CF VAR	.85263	.72282	.80462	.74243	.86780	.91429
VAR	.00032	.00182	.00081	.00426	.00373	.00026
RECSUM	.23177	.64954	.38815	.96760	.77388	.19408

ANALYSIS WITH .BSS/TSS(I)

	13LOANTM	14NECESS	15EQUITY	16CRERAT
1	.002961	.015418	.005472	.589449
2	.011335	.004245	.022545	.062335
3	.022999	.035403	.012953	.147642
4	.076923	.084639	.040000	0.000000
5	.005829	.016679	.010833	.036727
6	.084175	.164983	.084175	0.000000
7	.001826	.004676	.005131	.003210
8	.005102	0.000009	0.000000	.035714
9	0.000006	0.000000	0.000000	0.000000
10	.006536	.003811	.004138	0.000000
11	.035088	.122807	.017544	0.000000
SUBSUM	.25167	.45186	.20279	.87508
TRIALS	11	11	11	6
MEANBS	.02288	.04106	.01844	.14585
S D	.02893	.05454	.02359	.20336
CF VAR	1.26435	1.32772	1.27963	1.39437
VAR	.00084	.00297	.00056	.04136
RECSUM	.25167	.45186	.20279	1.60431

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 11 GRAND MEAN = .466925 ROUND CRIT. = .010TSST = 55.7

		SUB(VARIANT) SUMBSS/TSS(I)	RECONSTRUCTED SUMBSS/TSS(I)	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	AGE	.28709	.28709	.03317	1.27881	.00116	61.34825	11	9.0
2	MARITL	.23422	.23422	.03098	1.45477	.00096	58.04371	11	11.0
3	LOCAL	.07379	.07379	.00956	1.42561	.00009	15.76587	11	16.0
4	PHONE	.41953	.46149	.04524	1.07831	.00205	98.60275	10	6.0
5	JOBTIM	.50427	.58627	.04470	.97517	.00200	187.74353	11	5.0
6	RESTIM	.21192	.21192	.01454	.75476	.00021	45.27945	11	13.0
7	RESCAT	.23177	.23177	.01796	.85263	.00032	49.52878	11	12.0
8	TYPEMP	.59049	.64954	.04260	.72282	.00182	138.78252	10	4.0
9	INCOME	.38815	.38815	.02839	.80462	.00081	82.93387	11	8.0
10	BANKAC	.87963	.96760	.06531	.74243	.00426	206.74024	10	2.0
11	FINREF	.70352	.77388	.06105	.86788	.00373	165.34919	10	3.0
12	LOANAM	.19408	.19408	.01613	.91429	.00026	41.46764	11	15.0
13	LOANTM	.25167	.25167	.02893	1.26435	.00084	53.77376	11	10.0
14	NECESS	.45186	.45186	.05454	1.32772	.00297	96.54633	11	7.0
15	EQUITY	.20279	.20279	.02359	1.27963	.00056	43.32831	11	14.0
16	CRERAT	.87508	1.60431	.20336	1.39437	.04136	342.78239	6	1.0

```
GROUP  1         MEAN=      .46
N=   224         S.D.=      .50
       LEVEL 1

GROUP  3         MEAN=      .86 RSQ =  .589
N=   188         S.D.=      .35 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  2  3

GROUP  5         MEAN=      .92 RSQ =  .630
N=    93         S.D.=      .26 PROB= 0.000
PREDICTOR 11 FINREF
CODES  1  2

GROUP 11*FINAL MEAN=        .95 RSQ =  .707
N=    85         S.D.=      .21 PROB=  .001
PREDICTOR  4 PHONE
CODES  1

GROUP 10*FINAL MEAN=        .63 RSQ =  .707
N=     8         S.D.=      .48 PROB=  .001
PREDICTOR  4 PHONE
CODES  0

GROUP  4*FINAL MEAN=        .47 RSQ =  .630
N=    15         S.D.=      .50 PROB= 0.000
PREDICTOR 11 FINREF
CODES  0

GROUP  2         MEAN=      .09 RSQ =  .589
N=   116         S.D.=      .29 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  0  1

GROUP  7         MEAN=      .25 RSQ =  .634
N=    40         S.D.=      .43 PROB=  .000
PREDICTOR 16 CRERAT
CODES  1

GROUP  9         MEAN=      .45 RSQ =  .693
N=    20         S.D.=      .50 PROB= 0.000
PREDICTOR 10 BANKAC
CODES  2  1  3

GROUP  8*FINAL MEAN=        .05 RSQ =  .693
N=    20         S.D.=      .22 PROB= 0.000
PREDICTOR 10 BANKAC
CODES  0

GROUP  6*FINAL MEAN=        .01 RSQ =  .664
N=    76         S.D.=      .11 PROB=  .000
PREDICTOR 16 CRERAT
CODES  0
```

```
**************************************************
*GROUP   13*FINAL MEAN=        .64 RSQ =   .723*
* N=     11        S.D.=        .48 PROB=   .061*
* PREDICTOR   8 TYPEMP                          *
* CODES   1                                     *
*                                               *
*                                               *
**************************************************
                    *
                    *
                    *
                    *
                    *
                    *
******************************************
*                                        *
*                                        *
*GROUP    9        MEAN=        .45      *
* N=     20        S.D.=        .50      *
*           LEVEL 2                      *
*                                        *
*                                        *
******************************************
                    *
                    *
                    *
                    *
                    *
                    *
**************************************************
*GROUP   12*FINAL MEAN=        .22 RSQ =   .723*
* N=      9        S.D.=        .42 PROB=   .051*
* PREDICTOR   8 TYPEMP                          *
* CODES   0  2  3                               *
*                                               *
*                                               *
**************************************************
```

Exhibit 11 -- AID4UT OUTPUT FOR JUDGE #4

```
GROUP    N    TOTAL WEIGHT    SUM OF Y     SUM Y-SQUARE      T S S         MEAN        STD. DEV.
  1     224   224.00000       118.00000    118.00000       55.839286    .52678571    .49928201

                                          SPLIT SUMMARY

SPLIT GROUP   1 ON PREDICTOR  16   CRERAT           INTO GROUP   2 WITH CODES  0
                                                    AND GROUP    3 WITH CODES  1  2  3
      GROUP  2  N=    76 MEAN=      .04      GROUP   3  N=   148 MEAN=    .78    T=  14.59    RSQ=  .48919 F-RSQ=  212.6
SPLIT GROUP   3 ON PREDICTOR  16   CRERAT           INTO GROUP   4 WITH CODES  1
                                                    AND GROUP    5 WITH CODES  2  3
      GROUP  4  N=    40 MEAN=      .52      GROUP   5  N=   108 MEAN=    .87    T=  4.789    RSQ=  .55154 F-RSQ=  30.73
SPLIT GROUP   5 ON PREDICTOR  11   FINREF           INTO GROUP   6 WITH CODES  0
                                                    AND GROUP    7 WITH CODES  1  2
      GROUP  6  N=    15 MEAN=      .53      GROUP   7  N=    93 MEAN=    .92    T=  4.533    RSQ=  .58697 F-RSQ=  18.88
SPLIT GROUP   4 ON PREDICTOR  13   LOANTM           INTO GROUP   8 WITH CODES  0  1
                                                    AND GROUP    9 WITH CODES  2
      GROUP  8  N=    27 MEAN=      .70      GROUP   9  N=    13 MEAN=    .15    T=  3.711    RSQ=  .63449 F-RSQ=  28.47
SPLIT GROUP   7 ON PREDICTOR   4   PHONE            INTO GROUP  10 WITH CODES  0
                                                    AND GROUP   11 WITH CODES  1
      GROUP 10  N=     8 MEAN=      .50      GROUP  11  N=    85 MEAN=    .96    T=  5.418    RSQ=  .66276 F-RSQ=  18.26
SPLIT GROUP   8 ON PREDICTOR   9   INCOME           INTO GROUP  12 WITH CODES  0
                                                    AND GROUP   13 WITH CODES  1  2  3
      GROUP 12  N=     9 MEAN=      .44      GROUP  13  N=    18 MEAN=    .83    T=  2.192    RSQ=  .67902 F-RSQ=  10.99
SPLIT GROUP   6 ON PREDICTOR   8   TYPEMP           INTO GROUP  14 WITH CODES  3  1
                                                    AND GROUP   15 WITH CODES  2
      GROUP 14  N=     8 MEAN=      .25      GROUP  15  N=     7 MEAN=    .86    T=  2.755    RSQ=  .70366 F-RSQ=  17.96
SPLIT GROUP  13 ON PREDICTOR  11   FINREF           INTO GROUP  16 WITH CODES  0  1
                                                    AND GROUP   17 WITH CODES  2
      GROUP 16  N=     7 MEAN=      .57      GROUP  17  N=    11 MEAN=   1.00    T=  2.788    RSQ=  .71773 F-RSQ=  10.72

                                          FINAL SUMMARY

NCF   TOTAL TSS     TOTAL BSS     TOTAL WSS    R-SQUARED      R                      F-ANOVA    DF1   DF2
 9    55.839286     40.077631     15.761655    .71773170    .84719                   68.3359     8    215
```

ANALYSIS WITH . BSS/TSS(I)

TRIAL	1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	.009641	.001062	.006583	.037590	.079314	.012307
2	.034792	.017406	.000094	.077341	.051207	.014084
3	.000308	.004479	.010682	.104110	.074301	.002462
4	.010025	.060508	.029993	.007555	.087519	.097550
5	.001933	.005772	.634021	.243932	.112793	.014366
6	.026161	.082573	.108707	.009188	.008224	.005263
7	.082976	0.000000	.002976	0.000000	.035714	.035714
8	.005646	.030495	.018840	0.000000	.087805	.031723
9	.035082	.006955	.035082	.009280	.053549	.094699
10	.100000	.100000	.022222	.003077	.002597	.010000
11	.001082	0.000000	0.000000	0.000000	0.000000	.001082
SUBSUM	.22772	.30925	.26918	.48507	.51302	.23005
TRIALS	11	11	11	10	11	11
MEANBS	.02070	.02811	.02447	.04951	.04664	.02091
S D	.02804	.03454	.02943	.07401	.03719	.02560
CF VAR	1.35464	1.22850	1.20289	1.52582	.79738	1.22416
VAR	.00079	.00119	.00087	.00540	.00138	.00066
RECSUM	.22772	.30925	.26918	.53358	.51302	.23005

ANALYSIS WITH BSS/TSS(I)

TRIAL	7RESCAT	8TYPEMP	9INCOME	10BANKAC	11FINREF	12LOANAM
1	.049973	.053610	.021711	.091708	.076996	.018883
2	.073512	.063500	.053483	.096671	.112788	.001620
3	.061381	.085787	.018588	.025178	.162388	.001520
4	.014202	.023365	.050839	.091966	.062657	.003363
5	.083165	.053156	.049285	.042935	.025184	.000138
6	0.000000	.056861	.161184	.159211	.159211	.008224
7	0.000000	.368622	.241071	.247619	0.000000	.215561
8	.039262	.012776	.018944	.016396	.004065	.013572
9	.108872	.025805	.011841	.043317	.010959	.004169
10	0.000009	.022222	0.000000	.150769	.314286	.002597
11	0.000000	.290909	.113636	.113636	.212121	0.000000
SUBSUM	.42237	1.05662	.75050	.87945	1.14265	.26935
TRIALS	11	11	11	11	10	11
MEANBS	.03840	.09606	.06823	.07995	.11427	.02449
S D	.03604	.11330	.07142	.04650	.09412	.06069
CF VAR	.93856	1.17953	1.04675	.58166	.82368	2.47856
VAR	.00130	.01284	.00510	.00216	.00886	.00368
RECSUM	.42237	1.05662	.75050	.87945	1.25692	.26935

ANALYSIS WITH · BSS/TSS(I)

	13LOANTM	14NECESS	15EQUITY	16CRERAT
1	.000493	.022281	.000303	.489186
2	.031698	.000981	.002549	.135782
3	.009726	.000744	.001868	.078048
4	.265971	.000636	.020886	0.000000
5	.005629	.001933	.001978	.075800
6	.117867	.023026	.083795	0.000008
7	.038584	0.000000	0.000008	.008929
8	.018946	.000062	.024555	.060595
9	.020145	.181996	.062894	0.000000
10	.100000	0.000098	.100000	0.000000
11	0.000000	.010227	0.000008	0.000000
SUBSUM	.60826	.24099	.23883	.84914
TRIALS	10	11	11	6
MEANBS	.06083	.02191	.02171	.14152
S D	.07799	.05133	.03424	.15983
CF VAR	1.28226	2.34286	1.57714	1.12938
VAR	.00608	.00263	.00117	.02555
RECSUM	.66909	.24099	.23883	1.55676

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 11 GRAND MEAN = .588979 ROUND CRIT. = .010TSST = 55.8

		SUB(VARIANT) SUMBSS/TSS(I)	RECONSTRUCTED SUMBSS/TSS(I)	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	AGE	.22772	.22772	.02884	1.35664	.00079	38.66328	11	16.0
2	MARITL	.38925	.38925	.03454	1.22050	.00119	52.58615	11	10.0
3	LOCAL	.26918	.26918	.02843	1.20280	.00087	45.70298	11	12.0
4	PHONE	.48587	.53358	.07481	1.52582	.00548	90.59468	10	7.0
5	JOBTIM	.51382	.51302	.03719	.79738	.00138	87.10398	11	8.0
6	RESTIM	.23005	.23005	.02560	1.22416	.00066	39.05422	11	15.0
7	RESCAT	.42237	.42237	.03604	.93856	.00130	71.71186	11	9.0
8	TYPEMP	1.05662	1.05662	.11330	1.17953	.01284	179.39865	11	3.0
9	INCOME	.75059	.75050	.07142	1.04675	.00510	127.42337	11	5.0
10	BANKAC	.87945	.87945	.04650	.58166	.00216	149.31719	11	4.0
11	FINREF	1.14265	1.25692	.09412	.82368	.00886	213.48645	10	2.0
12	LOANAM	.26935	.26935	.06069	2.47856	.00368	45.73214	11	11.0
13	LOANTM	.60826	.66909	.07799	1.28226	.00608	113.60093	10	6.0
14	NECESS	.24099	.24099	.05133	2.34286	.00263	40.91594	11	13.0
15	EQUITY	.23883	.23883	.03424	1.57714	.00117	40.54935	11	14.0
16	CRERAT	.84914	1.55676	.15983	1.12938	.02555	264.31424	6	1.0

```
GROUP  1          MEAN=     .53
 N=   224         S.D.=     .50
        LEVEL 1

GROUP  2*FINAL    MEAN=     .04 RSQ =  .489
 N=    76         S.D.=     .19 PROB= 0.000
 PREDICTOR 16 CRERAT
 CODES  0

GROUP  3          MEAN=     .78 RSQ =  .489
 N=   148         S.D.=     .42 PROB= 0.000
 PREDICTOR 16 CRERAT
 CODES  1  2  3

GROUP  5          MEAN=     .87 RSQ =  .552
 N=   108         S.D.=     .34 PROB= 0.000
 PREDICTOR 16 CRERAT
 CODES  2  3

GROUP  4          MEAN=     .52 RSQ =  .552
 N=    40         S.D.=     .50 PROB= 0.000
 PREDICTOR 16 CRERAT
 CODES  1

GROUP  7          MEAN=     .92 RSQ =  .587
 N=    93         S.D.=     .26 PROB=  .000
 PREDICTOR 11 FINREF
 CODES  1  2

GROUP  6          MEAN=     .53 RSQ =  .587
 N=    15         S.D.=     .50 PROB=  .000
 PREDICTOR 11 FINREF
 CODES  0

GROUP  8          MEAN=     .70 RSQ =  .634
 N=    27         [illegible]     .46 PROB= 0.000
 PREDICTOR 13 LOANTM
 CODES  0  1

GROUP  9*FINAL    MEAN=     .15 RSQ =  .634
 N=    13         S.D.=     .36 PROB= 0.000
 PREDICTOR 13 LOANTM
 CODES  2
```

```
GROUP  15=FINAL MEAN=     .86 RSQ =  .764
 N=    7        S.D.=     .35 PROB=  .000
 PREDICTOR  8 TYPEMP
 CODES  2

GROUP   6        MEAN=     .53
 N=    15        S.D.=     .58
        LEVEL 2

GROUP  14=FINAL MEAN=     .25 RSQ =  .704
 N=    8        S.D.=     .43 PROB=  .000
 PREDICTOR  8 TYPEMP
 CODES  3  1
```

```
GROUP  11=FINAL MEAN=     .96 RSQ =  .663
 N=    85       S.D.=     .18 PROB=  .000
 PREDICTOR  4 PHONE
 CODES  1

GROUP   7        MEAN=     .92
 N=    93        S.D.=     .26
        LEVEL 2

GROUP  10=FINAL MEAN=     .50 RSQ =  .663
 N=    8        S.D.=     .50 PROB=  .000
 PREDICTOR  4 PHONE
 CODES  0
```

```
GROUP 1        MEAN=    .53
N=  224        S.D.=    .58
    LEVEL 1
```

```
GROUP 3        MEAN=    .78 RSQ =  .489
N=  148        S.D.=    .42 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 1 2 3
```

```
GROUP 2*FINAL MEAN=     .04 RSQ =  .489
N=   76        S.D.=    .19 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 0
```

```
GROUP 5        MEAN=    .87 RSQ =  .552
N=  188        S.D.=    .34 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 2 3
```

```
GROUP 4        MEAN=    .52 RSQ =  .552
N=   40        S.D.=    .50 PROB= 0.000
PREDICTOR 16 CRERAT
CODES 1
```

```
GROUP 7        MEAN=    .92 RSQ =  .567
N=   93        S.D.=    .26 PROB=  .000
PREDICTOR 11 FINREF
CODES 1 2
```

```
GROUP 6        MEAN=    .53 RSQ =  .587
N=   15        S.D.=    .50 PROB=  .000
PREDICTOR 11 FINREF
CODES 0
```

```
GROUP 8        MEAN=    .70 RSQ =  .634
N=   27        S.D.=    .46 PROB= 0.000
PREDICTOR 13 LOANTM
CODES 0 1
```

```
GROUP 9*FINAL MEAN=     .15 RSQ =  .634
N=   13        S.D.=    .36 PROB= 0.000
PREDICTOR 13 LOANTM
CODES 2
```

```
GROUP  15 FINAL MEAN=     .86 RSQ =  .704
 N=    7      S.D.=       .35 PROB=  .000
 PREDICTOR  8 TYPEMP
 CODES  2
```

```
GROUP   6        MEAN=     .53
 N=    15        S.D.=     .50
        LEVEL 2
```

```
GROUP  14 FINAL MEAN=     .25 RSQ =  .704
 N=    6      S.D.=       .43 PROB=  .000
 PREDICTOR  8 TYPEMP
 CODES  3  1
```

```
GROUP  11 FINAL MEAN=     .96 RSQ =  .663
 N=    85     S.D.=       .18 PROB=  .000
 PREDICTOR  4 PHONE
 CODES  1
```

```
GROUP   7        MEAN=     .92
 N=    93        S.D.=     .26
        LEVEL 2
```

```
GROUP  10 FINAL MEAN=     .50 RSQ =  .663
 N=    8      S.D.=       .50 PROB=  .000
 PREDICTOR  4 PHONE
 CODES  0
```

```
GROUP  17*FINAL MEAN=     1.00 RSQ =  .718
 N=     11          S.D.=     0.00 PROB=  .001
 PREDICTOR 11 FINREF
 CODES  2

GROUP  13        MEAN=     .83 RSQ =  .679
 N=     18       S.D.=     .37 PROB=  .001
 PREDICTOR  9 INCOME
 CODES  1  2  3

GROUP  16*FINAL MEAN=     .57 RSQ =  .718
 N=      7          S.D.=     .49 PROB=  .001
 PREDICTOR 11 FINREF
 CODES  0  1

GROUP   8        MEAN=     .70
 N=     27       S.D.=     .46
       LEVEL 2

GROUP  12*FINAL MEAN=     .44 RSQ =  .679
 N=      9       S.D.=     .50 PROB=  .061
 PREDICTOR  9 INCOME
 CODES  0
```

Exhibit 12 -- AID4UT OUTPUT FOR JUDGE #5

```
GROUP    N    TOTAL WEIGHT     SUM OF Y      SUM Y-SQUARE        T S S          MEAN         STD. DEV.
  1     224   224.00000       111.00000      111.00000        55.995536      .49553571     .49998007


                                                  SPLIT SUMMARY

SPLIT GROUP   1 ON PREDICTOR  16   CRERAT             INTO GROUP   2 WITH CODES  0  1
                                                      AND GROUP    3 WITH CODES  2  3
      GROUP   2  N=   116 MEAN=    .21      GROUP   3  N=  108 MEAN=     .81    T=  11.13     RSQ=  .35796 F-RSQ=  123.8
*************************************************************************************************************************
SPLIT GROUP   2 ON PREDICTOR  16   CRERAT             INTO GROUP   4 WITH CODES  0
                                                      AND GROUP    5 WITH CODES  1
      GROUP   4  N=    76 MEAN=    .08      GROUP   5  N=   40 MEAN=     .45    T=  5.164     RSQ=  .42240 F-RSQ=  24.65
*************************************************************************************************************************
SPLIT GROUP   3 ON PREDICTOR  11   FINREF             INTO GROUP   6 WITH CODES  0  1
                                                      AND GROUP    7 WITH CODES  2
      GROUP   6  N=    34 MEAN=    .50      GROUP   7  N=   74 MEAN=     .95    T=  6.323     RSQ=  .50514 F-RSQ=  36.78
*************************************************************************************************************************
SPLIT GROUP   5 ON PREDICTOR  10   BANKAC             INTO GROUP   8 WITH CODES  0
                                                      AND GROUP    9 WITH CODES  3  2  1
      GROUP   8  N=    20 MEAN=    .20      GROUP   9  N=   20 MEAN=     .70    T=  3.583     RSQ=  .54978 F-RSQ=  21.72
*************************************************************************************************************************
SPLIT GROUP   6 ON PREDICTOR   7   RESC-T             INTO GROUP  10 WITH CODES  1
                                                      AND GROUP   11 WITH CODES  2  3
      GROUP  10  N=    23 MEAN=    .35      GROUP  11  N=   11 MEAN=     .82    T=  2.772     RSQ=  .57918 F-RSQ=  15.23
*************************************************************************************************************************
SPLIT GROUP   4 ON PREDICTOR   5   JOBTIM             INTO GROUP  12 WITH CODES  0  1
                                                      AND GROUP   13 WITH CODES  2
      GROUP  12  N=    60 MEAN=    .02      GROUP  13  N=   16 MEAN=     .31    T=  4.302     RSQ=  .59893 F-RSQ=  10.68
*************************************************************************************************************************
SPLIT GROUP  10 ON PREDICTOR   1   AGE                INTO GROUP  14 WITH CODES  0  1
                                                      AND GROUP   15 WITH CODES  2  3  4
      GROUP  14  N=     8 MEAN=    .63      GROUP  15  N=   15 MEAN=     .20    T=  2.152     RSQ=  .61576 F-RSQ=  9.461
*************************************************************************************************************************
SPLIT GROUP   9 ON PREDICTOR  13   LOANTM             INTO GROUP  16 WITH CODES  0  1
                                                      AND GROUP   17 WITH CODES  2
      GROUP  16  N=    15 MEAN=    .93      GROUP  17  N=    5 MEAN=     .40    T=  1.732     RSQ=  .62647 F-RSQ=  6.168
*************************************************************************************************************************
SPLIT GROUP   7 ON PREDICTOR   5   JOBTIM             INTO GROUP  18 WITH CODES  0
                                                      AND GROUP   19 WITH CODES  1  2
      GROUP  18  N=    20 MEAN=    .80      GROUP  19  N=   54 MEAN=    1.00    T=  3.624     RSQ=  .63690 F-RSQ=  6.144
*************************************************************************************************************************
SPLIT GROUP  13 ON PREDICTOR  11   FINREF             INTO GROUP  20 WITH CODES  0  1
                                                      AND GROUP   21 WITH CODES  2
      GROUP  20  N=     9 MEAN=   0.00      GROUP  21  N=    7 MEAN=     .71    T=  4.437     RSQ=  .67277 F-RSQ=  23.35
*************************************************************************************************************************
SPLIT GROUP  15 ON PREDICTOR   6   RESTIM             INTO GROUP  22 WITH CODES  0  1
                                                      AND GROUP   23 WITH CODES  2
      GROUP  22  N=    10 MEAN=   0.00      GROUP  23  N=    5 MEAN=     .60    T=  3.606     RSQ=  .69420 F-RSQ=  14.85
*************************************************************************************************************************


*************************************************************************************************************************

                                                  FINAL SUMMARY


NCF     TOTAL TSS       TOTAL BSS       TOTAL WSS      R-SQUARED       R                          F-ANOVA   DF1   DF2

 12     55.995536       38.872267       17.123268      .69420297     .83319                       43.7518    11   212

*************************************************************************************************************************
```

ANALYSIS WITH BSS/TSS(1)

TRIAL	1AGE	2MARITL	3LOCAL	4PHONE	5JOBTIM	6RESTIM
1	.016327	.005405	.014302	.046182	.123979	.039573
2	.012475	.008569	.002404	.003682	.082277	.014105
3	.006288	.007258	.049154	.089201	.073688	.047050
4	.090909	.058719	.032079	.011431	.050783	.044995
5	.035573	.062069	.087719	.062069	.019231	.179467
6	.029793	.014505	.029793	.019355	.200040	.114307
7	.180625	0.000000	.111364	0.000000	.051062	.096429
8	.142857	.042386	.004329	0.000000	.042386	.001289
9	.006717	.011932	.033551	.134594	.154286	.023588
10	.059394	0.000000	.059394	0.000000	0.000000	.026777
11	.047619	.083333	.023438	0.000000	0.000000	.010417
12	.047619	.134615	0.000000	.024725	.002976	.040404
13	.166667	0.000000	.004630	0.000000	.074074	.040179
14	.074074	0.000000	.285714	0.000000	.004630	.500000
15	.185185	0.000000	0.000000	0.000000	.001852	.266667
16	.020716	.003390	.015856	.004682	.042871	.067797
SUBSUM	1.12284	.43220	.75373	.39593	.93413	1.59322
TRIALS	16	15	16	16	14	16
MEANBS	.07018	.02881	.04711	.02475	.06672	.09958
S D	.06185	.03850	.06904	.03825	.05662	.12219
CF VAR	.88136	1.33605	1.46553	1.54585	.84863	1.22713
VAR	.00383	.00148	.00477	.00146	.00321	.01493
RECSUM	1.12284	.46102	.75373	.39593	1.06758	1.59322

ANALYSIS WITH BSS/TSS(1)

TRIAL	7RESCAT	8TYPEMP	9INCOME	10BANFAC	11FINREF	12LOANAM
1	.002749	.049916	.034536	.151013	.160080	.013952
2	.013778	.004391	.004623	.061017	.065217	.092502
3	.131279	.078378	.042276	.133829	.273865	.002530
4	.033472	.000421	.058719	.252525	.090909	.008264
5	.193676	.125000	.132275	.171930	.087719	0.000000
6	.059291	.014192	.032407	.041191	.085714	.022754
7	0.000000	.054444	.081027	.111364	.001010	.001010
8	.015873	.015873	0.000000	0.000000	.142857	.097222
9	.063673	.051455	.032827	.021084	0.000000	.004638
10	0.000000	.018182	.104185	.059394	.584416	0.000000
11	0.000000	.127604	.083333	0.000000	0.000000	.010989
12	.083333	.083333	.062500	0.000000	.107143	0.000000
13	0.000000	.074074	0.000000	0.000000	0.000000	.166667
14	0.000000	0.000000	.004630	.074074	.004630	.004630
15	0.000000	.185185	0.000000	0.000000	0.000000	0.000000
16	.039548	.036574	.015856	.020716	.015856	.014831
SUBSUM	.71667	.91901	.69519	1.10053	1.62742	.34999
TRIALS	14	16	16	15	14	16
MEANBS	.05119	.05744	.04345	.07337	.11624	.02187
S D	.05513	.05138	.03949	.07366	.15008	.04384
CF VAR	1.07688	.89446	.90893	1.00395	1.29108	2.00421
VAR	.00304	.00264	.00156	.00543	.02252	.00192
RECSUM	.81906	.91901	.69519	1.17390	1.85991	.34999

ANALYSIS WITH BSS/TSS(I)

	13LOANAM	14NECESS	15EQUITY	16CRERAT
1	.066981	.021474	.001264	.357964
2	.035580	.039245	.001556	.189559
3	.036187	.000094	.004030	.145261
4	.093499	.013964	.030303	0.000000
5	.076923	.035573	.005291	.136364
6	.032727	.056880	.006036	0.000000
7	.007756	.007440	.026759	.158170
8	.142857	.015873	.015873	0.000000
9	.003253	.027429	.008816	.047250
10	.096182	0.000000	.016198	0.000000
11	.040404	0.000000	0.000000	.040404
12	.107143	.047619	.083333	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	.074074	0.000000	0.000000	0.000000
15	.001852	0.000000	0.000000	.001852
16	.008475	.075501	.004688	0.000000
SUBSUM	.76581	.34109	.20415	1.07682
TRIALS	16	16	16	9
MEANBS	.04786	.02132	.01276	.11965
S D	.04365	.02299	.02042	.10740
CF VAR	.91207	1.07819	1.60054	.89765
VAR	.00191	.00053	.00042	.01153
RECSUM	.76581	.34109	.20415	1.91435

SUMMARY TABLE 16 VARIABLES, NO OF SUBGROUPS IS 16 GRAND MEAN = .982299 ROUND CRIT. = .0161SST = 56.0

		SUB(VARIANT) SUMBSS/TSS(I)	RECONSTRUCTED SUMBSS/TSS(I)	S D	COEFF VAR	VARIANCE PCT	GRO MEAN	N GROUPS	RANK
1	AGE	1.12284	1.12284	.06185	.88136	.00383	124.44192	16	5.0
2	MARITL	.43229	.46102	.03850	1.33605	.00148	51.09353	15	12.0
3	LOCAL	.75373	.75373	.06984	1.46553	.00477	83.53393	16	10.0
4	PHONE	.39593	.39593	.03825	1.54585	.00146	43.87974	16	13.0
5	JOBTIM	.93413	1.06758	.05662	.84863	.00321	118.31806	14	6.0
6	RESTIM	1.59322	1.59322	.12219	1.22713	.01493	176.57374	16	3.0
7	RESCAT	.71667	.81906	.05513	1.07688	.00304	90.77426	14	8.0
8	TYPEMP	.91901	.91901	.05138	.89446	.00264	101.85258	16	7.0
9	INCOME	.69519	.69519	.03949	.90893	.00156	77.04691	16	11.0
10	BANKAC	1.10053	1.17390	.07366	1.00395	.00543	130.10054	15	4.0
11	FINREF	1.62742	1.85991	.15008	1.29108	.02252	206.13054	14	2.0
12	LOANAM	.34999	.34999	.04384	2.00421	.00192	38.78865	16	14.0
13	LOANTM	.76581	.76581	.04365	.91207	.00191	84.87360	16	9.0
14	NECESS	.34109	.34109	.02299	1.07819	.00053	37.80257	16	15.0
15	EQUITY	.20415	.20415	.02042	1.60054	.00042	22.62530	16	16.0
16	CRERAT	1.07682	1.91435	.10740	.89765	.01153	212.16411	9	1.0

```
GROUP  1         MEAN=    .50
N=   224         S.D.=    .50
     LEVEL 1

GROUP  2         MEAN=    .21 RSQ =  .358
N=   116         S.D.=    .41 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  0  1

GROUP  3         MEAN=    .81 RSQ =  .358
N=   108         S.D.=    .40 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  2  3

GROUP  7         MEAN=    .95 RSQ =  .505
N=    74         S.D.=    .23 PROB= 0.000
PREDICTOR 11 FINREF
CODES  2

GROUP  6         MEAN=    .50 RSQ =  .505
N=    34         S.D.=    .50 PROB= 0.000
PREDICTOR 11 FINREF
CODES  0  1

GROUP  5         MEAN=    .43 RSQ =  .422
N=    40         S.D.=    .50 PROB= 0.000
PREDICTOR 16 CRERA
CODES  1

GROUP  4         MEAN=    .08 RSQ =  .422
N=    76         S.D.=    .27 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  0

GROUP 19*FINAL MEAN=     1.00 RSQ =  .637
N=    54         S.D.=   0.00 PROB=  .014
PREDICTOR  5 JOBTIM
CODES  1  2

GROUP 18*FINAL MEAN=      .80 RSQ =  .637
N=    20         S.D.=    .40 PROB=  .014
PREDICTOR  5 JOBTIM
CODES  0

GROUP 11*FINAL MEAN=      .82 RSQ =  .579
N=    11         S.D.=    .39 PROB=  .000
PREDICTOR  7 RESCAT
CODES  2  3

GROUP 10         MEAN=    .35 RSQ =  .579
N=    23         S.D.=    .48 PROB=  .000
PREDICTOR  7 RESCAT
CODES  1

GROUP  9         MEAN=    .70 RSQ =  .550
N=    20         S.D.=    .46 PROB= 0.000
PREDICTOR 10 BANKAC
CODES  3  2  1

GROUP  8*FINAL MEAN=      .20 RSQ =  .550
N=    20         S.D.=    .40 PROB= 0.000
PREDICTOR 10 BANKAC
CODES  0

GROUP 13         MEAN=    .31 RSQ =  .599
N=    16         S.D.=    .46 PROB=  .001
PREDICTOR  5 JOBTIM
CODES  2

GROUP 12*FINAL MEAN=      .02 RSQ =  .599
N=    60         S.D.=    .13 PROB=  .001
PREDICTOR  5 JOBTIM
CODES  0  1
```

```
GROUP  14*FINAL MEAN=      .63 RSQ =  .616
 N=     8        S.D.=     .48 PROB=  .002
 PREDICTOR  1 AGE
 CODES  0  1

GROUP  10        MEAN=     .35
 N=    23        S.D.=     .48
          LEVEL 2

GROUP  15        MEAN=     .20 RSQ =  .616
 N=    15        S.D.=     .40 PROB=  .002
 PREDICTOR  1 AGE
 CODES  2  3  4

GROUP  23*FINAL MEAN=      .60 RSQ =  .694
 N=     5        S.D.=     .49 PROB=  .588
 PREDICTOR  6 RESTIM
 CODES  2

GROUP  22*FINAL MEAN=     0.00 RSQ =  .694
 N=    10        S.D.=    0.00 PROB=  .000
 PREDICTOR  6 RESTIM
 CODES  0  1
```

```
GROUP  16*FINAL MEAN=      .80 RSQ =  .626
 N=    15      S.D.=       .40 PROB=  .014
 PREDICTOR 13 LOANTM
 CODES  0  1

GROUP   9      MEAN=       .70
 N=    28      S.D.=       .46
     LEVEL 2

GROUP  17*FINAL MEAN=      .40 RSQ =  .626
 N=     5      S.D.=       .49 PROB=  .014
 PREDICTOR 13 LOANTM
 CODES  2
```

```
GROUP  21*FINAL MEAN=      .71 RSQ =  .673
 N=     7      S.D.=       .45 PROB= 0.000
 PREDICTOR 11 FINREF
 CODES  2

GROUP  13      MEAN=       .31
 N=    16      S.D.=       .45
     LEVEL 2

GROUP  20*FINAL MEAN=     0.00 RSQ =  .673
 N=     9      S.D.=      0.00 PROB= 0.000
 PREDICTOR 11 FINREF
 CODES  0  1
```

Exhibit 13 -- AID4UT OUTPUT FOR 4/5 VOTE

```
GROUP    N     TOTAL WEIGHT      SUM OF Y       SUM Y-SQUARE        T S S            MEAN          STD. DEV.
  1     224    224.00000        97.000000       97.000000         54.995536       .43303571       .49549549

                                                   SPLIT SUMMARY

SPLIT GROUP   1 ON PREDICTOR  16   CRERAT            INTO GROUP   2 WITH CODES  0  1
                                                     AND GROUP    3 WITH CODES  2  3
      GROUP   2  N=   116 MEAN=      .89       GROUP   3  N=   108 MEAN=      .80    T=  14.91     RSQ=  .50841 F-RSQ=  222.4
**********************************************************************************************************************
SPLIT GROUP   3 ON PREDICTOR  11   FINREF            INTO GROUP   4 WITH CODES  0
                                                     AND GROUP    5 WITH CODES  1  2
      GROUP   4  N=    15 MEAN=      .20       GROUP   5  N=    93 MEAN=      .89    T=  7.614     RSQ=  .61303 F-RSQ=  64.32
**********************************************************************************************************************
SPLIT GROUP   2 ON PREDICTOR  16   CRERAT            INTO GROUP   6 WITH CODES  0
                                                     AND GROUP    7 WITH CODES  1
      GROUP   6  N=    76 MEAN=     0.00       GROUP   7  N=    40 MEAN=      .27    T=  5.323     RSQ=  .64907 F-RSQ=  22.59
**********************************************************************************************************************
SPLIT GROUP   5 ON PREDICTOR   4   PHONE             INTO GROUP   8 WITH CODES  0
                                                     AND GROUP    9 WITH CODES  1
      GROUP   8  N=     8 MEAN=      .50       GROUP   9  N=    85 MEAN=      .93    T=  4.024     RSQ=  .67358 F-RSQ=  16.45
**********************************************************************************************************************
SPLIT GROUP   7 ON PREDICTOR  10   BANKAC            INTO GROUP  10 WITH CODES  1  0
                                                     AND GROUP   11 WITH CODES  3  2
      GROUP  10  N=    21 MEAN=      .05       GROUP  11  N=    19 MEAN=      .53    T=  3.907     RSQ=  .71515 F-RSQ=  31.81
**********************************************************************************************************************
SPLIT GROUP   9 ON PREDICTOR   6   RESTIM            INTO GROUP  12 WITH CODES  0
                                                     AND GROUP   13 WITH CODES  1  2
      GROUP  12  N=    22 MEAN=      .77       GROUP  13  N=    63 MEAN=      .98    T=  3.532     RSQ=  .72840 F-RSQ=  10.59
**********************************************************************************************************************
SPLIT GROUP  11 ON PREDICTOR  11   FINREF            INTO GROUP  14 WITH CODES  0  1
                                                     AND GROUP   15 WITH CODES  2
      GROUP  14  N=     5 MEAN=      .20       GROUP  15  N=    14 MEAN=      .64    T=  1.749     RSQ=  .74153 F-RSQ=  10.98
**********************************************************************************************************************
SPLIT GROUP  12 ON PREDICTOR  11   FINREF            INTO GROUP  16 WITH CODES  1
                                                     AND GROUP   17 WITH CODES  2
      GROUP  16  N=     6 MEAN=      .33       GROUP  17  N=    16 MEAN=      .94    T=  3.745     RSQ=  .77050 F-RSQ=  27.13
**********************************************************************************************************************
SPLIT GROUP   4 ON PREDICTOR   8   TYPEMP            INTO GROUP  18 WITH CODES  1  3
                                                     AND GROUP   19 WITH CODES  2
      GROUP  18  N=     6 MEAN=     0.00       GROUP  19  N=     7 MEAN=      .43    T=  2.280     RSQ=  .78297 F-RSQ=  12.29
**********************************************************************************************************************


**********************************************************************************************************************
                                                   FINAL SUMMARY

NCF    TOTAL TSS       TOTAL BSS        TOTAL WSS       R-SQUARED       R                      F-ANOVA   DF1   DF2

 10    54.995536       43.059623        11.935913       .78296579     .88485                   85.7800    9    214

**********************************************************************************************************************
```

```
*GROUP   1         MEAN=        .43
* N=   224         S.D.=        .56
        LEVEL 1

*GROUP   3         MEAN=        .80 RSQ =  .500*
* N=   108         S.D.=        .40 PROB= 0.000*
* PREDICTOR 16 CRERAT
* CODES  2  3

*GROUP   5         MEAN=        .89 RSQ =  .613*
* N=    93         S.D.=        .31 PROB= C.000*
* PREDICTOR 11 FINREF
* CODES  1  2

*GROUP   9         MEAN=        .93 RSQ =  .674*
* N=    85         S.D.=        .26 PROB=  .000*
* PREDICTOR  4 PHONE
* CODES  1

*GROUP   8*FINAL MEAN=          .50 RSQ =  .674*
* N=     8         S.D.=        .50 PROB=  .000*
* PREDICTOR  4 PHONE
* CODES  0

*GROUP   4         MEAN=        .20 RSQ =  .613*
* N=    15         S.D.=        .40 PROB= 0.000*
* PREDICTOR 11 FINREF
* CODES  0

*GROUP  19*FINAL MEAN=          .43 RSQ =  .783*
* N=     7         S.D.=        .49 PROB=  .001*
* PREDICTOR  6 TYPEMP
* CODES  2

*GROUP  18*FINAL MEAN=         0.00 RSQ =  .783*
* N=     8         S.D.=       0.00 PROB=  .061*
* PREDICTOH  5 TYPEMP
* CODES  1  3

*GROUP   2         MEAN=        .09 RSQ =  .500*
* N=   116         S.D.=        .29 PROB= 0.000*
* PREDICTOR 16 CRERAT
* CODES  0  1

*GROUP   7         MEAN=        .27 RSQ =  .649*
* N=    40         S.D.=        .45 PROB= C.000*
* PREDICTOH 16 CRERAT
* CODES  1

*GROUP  11         MEAN=        .53 RSQ =  .715*
* N=    15         S.D.=        .50 PROB= 0.000*
* PREDICTOR 16 BANKAC
* CODES  3  2

*GROUP  10*FINAL MEAN=          .85 RSQ =  .715*
* N=    21         S.D.=        .21 PROB= 0.000*
* PREDICTOR 10 BANKAC
* CODES  1  0

*GROUP   6*FINAL MEAN=         0.00 RSQ =  .649*
* N=    76         S.D.=       0.00 PROB= 0.000*
* PREDICTOR 16 CRERAT
* CODES  0
```

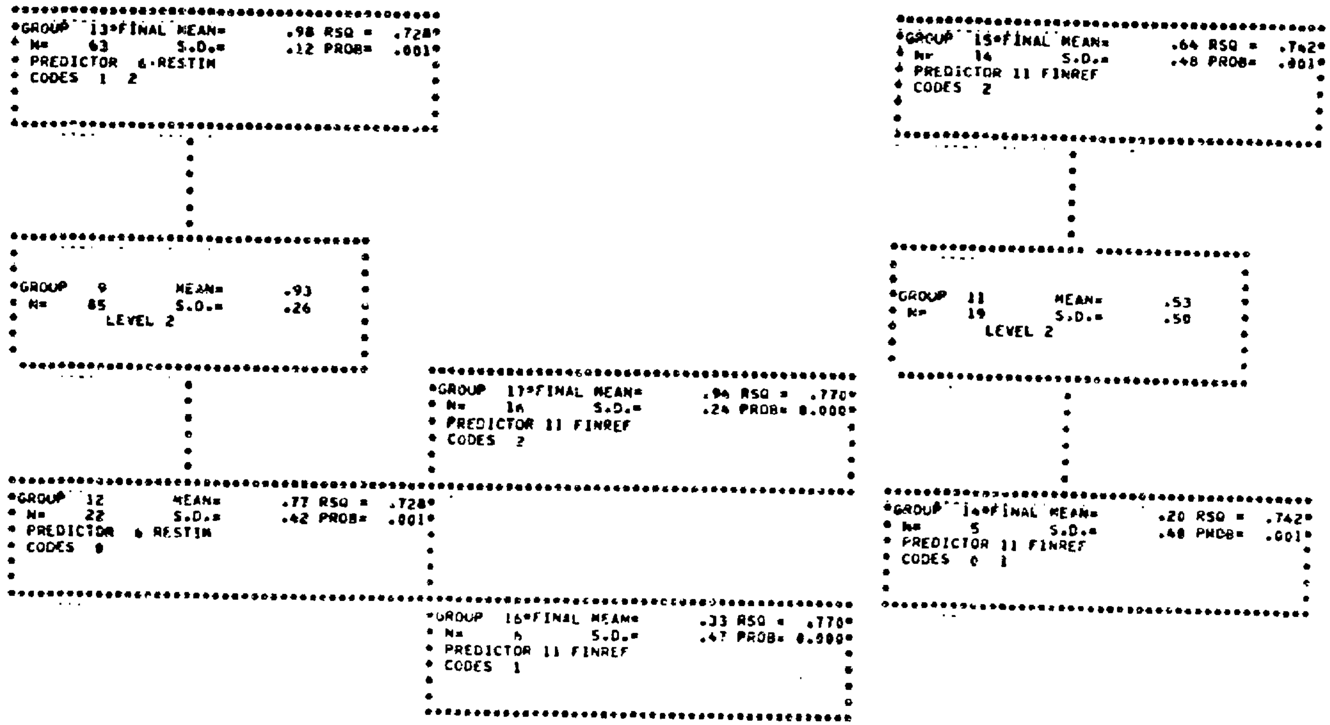

GROUP 13 FINAL MEAN= .98 RSQ = .728
N= 63 S.D.= .12 PROB= .001
PREDICTOR 6 RESTIN
CODES 1 2
GROUP 9 MEAN= .93
N= 85 S.D.= .26
LEVEL 2
GROUP 12 MEAN= .77 RSQ = .728
N= 22 S.D.= .42 PROB= .001
PREDICTOR 6 RESTIN
CODES 0
GROUP 17 FINAL MEAN= .94 RSQ = .770
N= 16 S.D.= .24 PROB= 0.000
PREDICTOR 11 FINREF
CODES 2
GROUP 16 FINAL MEAN= .33 RSQ = .770
N= 6 S.D.= .47 PROB= 0.000
PREDICTOR 11 FINREF
CODES 1
GROUP 15 FINAL MEAN= .64 RSQ = .742
N= 14 S.D.= .48 PROB= .001
PREDICTOR 11 FINREF
CODES 2
GROUP 11 MEAN= .53
N= 19 S.D.= .50
LEVEL 2
GROUP 14 FINAL MEAN= .20 RSQ = .742
N= 5 S.D.= .40 PROB= .001
PREDICTOR 11 FINREF
CODES 0 1

Exhibit 14 -- AID4UT TREE FOR JUDGE #6

```
GROUP   1        MEAN=     .71
N=    224        S.D.=     .45
         LEVEL 1
```

```
GROUP   3        MEAN=     .88 RSQ =  .269
N=    148        S.D.=     .33 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  1  2  3
```

```
GROUP   2        MEAN=     .38 RSQ =  .269
N=     76        S.D.=     .49 PROB= 0.000
PREDICTOR 16 CRERAT
CODES  0
```

```
GROUP   7*FINAL MEAN=      .96 RSQ =  .394
N=    105        S.D.=     .19 PROB=  .000
PREDICTOR  6 RESTIN
CODES  1  2
```

```
GROUP   6        MEAN=     .67 RSQ =  .394
N=     43        S.D.=     .47 PROB=  .000
PREDICTOR  6 RESTIN
CODES  0
```

```
GROUP   5        MEAN=     .70 RSQ =  .339
N=     23        S.D.=     .46 PROB= 0.000
PREDICTOR  6 RESTIN
CODES  2
```

```
GROUP   4        MEAN=     .25 RSQ =  .339
N=     53        S.D.=     .43 PROB= 0.000
PREDICTOR  6 RESTIN
CODES  0  1
```

```
GROUP  11        MEAN=     .82 RSQ =  .477
N=     28        S.D.=     .38 PROB=  .000
PREDICTOR 11 FINREF
CODES  2  0
```

```
GROUP  10        MEAN=     .40 RSQ =  .477
N=     15        S.D.=     .49 PROB=  .000
PREDICTOR 11 FINREF
CODES  1
```

```
GROUP  15*FINAL MEAN=     1.00 RSQ =  .540
N=     10        S.D.=    0.00 PROB=  .000
PREDICTOR  7 RESCAT
CODES  2
```

```
GROUP  14        MEAN=     .46 RSQ =  .540
N=     13        S.D.=     .50 PROB=  .000
PREDICTOR  7 RESCAT
CODES  3  1
```

```
GROUP   9        MEAN=     .64 RSQ =  .440
N=     11        S.D.=     .48 PROB=  .000
PREDICTOR  9 INCOME
CODES  3
```

```
GROUP   8        MEAN=     .14 RSQ =  .440
N=     42        S.D.=     .35 PROB=  .000
PREDICTOR  9 INCOME
CODES  0  1  2
```

```
GROUP 17*FINAL MEAN=   1.00 RSQ =  .581     GROUP 19*FINAL MEAN=   .80 RSQ =  .607     GROUP 21*FINAL MEAN=   .63 RSQ =  .619
N=    19      S.D.=    0.00 PROB= 0.000     N=     5      S.D.=    .40 PROB=  .000     N=     3      S.D.=    .48 PROB=  .010
PREDICTOR 16 CRERAT                         PREDICTOR 15 EQUITY                         PREDICTOR  9 INCOME
CODES  2  3                                 CODES  1  2                                 CODES  1  2  3

GROUP  11      MEAN=   .82                  GROUP  18      MEAN=   .40                  GROUP  14      MEAN=   .46
N=     28      S.D.=   .38                  N=     15      S.D.=   .49                  N=     13      S.D.=   .50
       LEVEL 2                                     LEVEL 2                                     LEVEL 2

GROUP 16*FINAL MEAN=   .44 RSQ =  .581      GROUP 18*FINAL MEAN=   .20 RSQ =  .607     GROUP 20*FINAL MEAN=   .20 RSQ =  .619
N=     9      S.D.=    .50 PROB= 0.000      N=    10      S.D.=    .40 PROB=  .000     N=     5      S.D.=    .40 PROB=  .010
PREDICTOR 16 CRERAT                         PREDICTOR 15 EQUITY                         PREDICTOR  9 INCOME
CODES  1                                    CODES  0                                    CODES  0
```

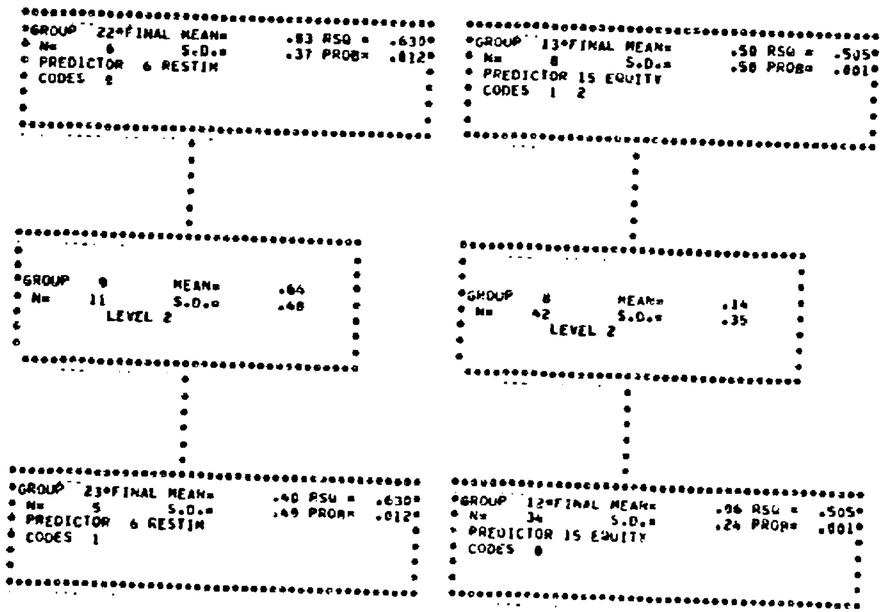

GROUP 22*FINAL MEAN= .83 RSQ = .630
N= 6 S.D.= .37 PROB= .012
PREDICTOR 6 RESTIM
CODES 2
GROUP 9 MEAN= .64
N= 11 S.D.= .48
LEVEL 2
GROUP 23*FINAL MEAN= .40 RSQ = .630
N= 5 S.D.= .49 PROB= .012
PREDICTOR 6 RESTIM
CODES 1
GROUP 13*FINAL MEAN= .50 RSQ = .505
N= 8 S.D.= .50 PROB= .001
PREDICTOR 15 EQUITY
CODES 1 2
GROUP 8 MEAN= .14
N= 42 S.D.= .35
LEVEL 2
GROUP 12*FINAL MEAN= .06 RSQ = .505
N= 34 S.D.= .24 PROB= .001
PREDICTOR 15 EQUITY
CODES 0

Exhibit 15 -- AID4UT TREE FOR JUDGE #7

```
GROUP   7 FINAL MEAN=     1.00 RSQ =  .547
 N=    42       S.D.=     0.00 PROB=  .000
 PREDICTOR 16 CRERAT
 CODES  3

GROUP   3       MEAN=      .68 RSQ =  .465
 N=   108       S.D.=      .33 PROB= 0.000
 PREDICTOR 16 CRERAT
 CODES  2  3

GROUP  11       MEAN=      .88 RSQ =  .015
 N=    26       S.D.=      .32 PROB=  .000
 PREDICTOR  9 INCOME
 CODES  2  3

GROUP   6       MEAN=      .72 RSQ =  .547
 N=    46       S.D.=      .45 PROB=  .000
 PREDICTOR 16 CRERAT
 CODES  2

GROUP  10       MEAN=      .50 RSQ =  .615
 N=    20       S.D.=      .50 PROB=  .000
 PREDICTOR  9 INCOME
 CODES  0  1

GROUP   1       MEAN=      .53
 N=   224       S.D.=      .50
        LEVEL 1

GROUP   8       MEAN=      .71 RSQ =  .585
 N=    14       S.D.=      .45 PROB=  .000
 PREDICTOR  3 LOCAL
 CODES  0

GROUP   5       MEAN=      .40 RSQ =  .509
 N=    40       S.D.=      .49 PROB=  .000
 PREDICTOR 16 CRERAT
 CODES  1

GROUP   9       MEAN=      .23 RSQ =  .585
 N=    26       S.D.=      .42 PROB=  .000
 PREDICTOR  3 LOCAL
 CODES  1

GROUP   2       MEAN=      .20 RSQ =  .465
 N=   116       S.D.=      .40 PROB= 0.000
 PREDICTOR 16 CRERAT
 CODES  0  1

GROUP  13       MEAN=      .23 RSQ =  .625
 N=    22       S.D.=      .42 PROB=  .016
 PREDICTOR  7 RESCAT
 CODES  2

GROUP   4       MEAN=      .09 RSQ =  .509
 N=    76       S.D.=      .29 PROB=  .000
 PREDICTOR 16 CRERAT
 CODES  0

GROUP  12 FINAL MEAN=      .04 RSQ =  .625
 N=    54       S.D.=      .19 PROB=  .016
 PREDICTOR  7 RESCAT
 CODES  0  3  1
```

```
GROUP 24 FINAL MEAN=  1.00 RSQ = .734
N=   19      S.D.=    0.00 PROB= .000
PREDICTOR 13 LOANTM
CODES 0 1

GROUP 1[illegible]   MEAN=  .88
N=   26      S.D.=  .32
        LEVEL 2

GROUP 25 FINAL MEAN=  .57 RSQ = .734
N=    7      S.D.=    .59 PROB= .000
PREDICTOR 13 LOANTM
CODES 2

GROUP 15 FINAL MEAN=  .86 RSQ = .650
N=    7      S.D.=    .35 PROB= .000
PREDICTOR 6 RESTIM
CODES 2

GROUP 10     MEAN=  .50
N=   20      S.D.=  .58
        LEVEL 2

GROUP 14     MEAN=  .31 RSQ = .650
N=   13      S.D.=  .46 PROB= .000
PREDICTOR 6 RESTIM
CODES 0 1

GROUP 23 FINAL MEAN=  .67 RSQ = .717
N=    6      S.D.=    .67 PROB= .000
PREDICTOR 11 FINREF
CODES 2

GROUP 22 FINAL MEAN=  0.00 RSQ = .717
N=    7      S.D.=    0.00 PROB= .000
PREDICTOR 11 FINREF
CODES 0 1
```

```
GROUP 21 FINAL MEAN=   1.00 RSQ = .691
N=    6       S.D.=    0.00 PROB= .001
PREDICTOR 6 RESTIM
CODES 1 2
```

```
GROUP 17 FINAL MEAN=   .50 RSQ = .665
N=    8       S.D.=    .50 PROB= .002
PREDICTOR 14 NECESS
CODES 1
```

```
GROUP  8      MEAN=    .71
N=    14      S.D.=    .45
     LEVEL 2
```

```
GROUP  9      MEAN=    .23
N=    26      S.D.=    .42
     LEVEL 2
```

```
GROUP 27 FINAL MEAN=   .40 RSQ = .744
N=    5       S.D.=    .49 PROB= .004
PREDICTOR 8 TYPEMP
CODES 2
```

```
GROUP 20 FINAL MEAN=   .50 RSQ = .691
N=    8       S.D.=    .50 PROB= .001
PREDICTOR 6 RESTIM
CODES 0
```

```
GROUP 16      MEAN=    .11 RSQ = .665
N=    18      S.D.=    . 1 PROB= .002
PREDICTOR 14 NECESS
CODES 0
```

```
GROUP 26 FINAL MEAN=   0.00 RSQ = .744
N=    13      S.D.=    0.00 PROB= .004
PREDICTOR 6 TYPEMP
CODES 1 3
```

```
*GROUP  18*FINAL MEAN=      .50 RSQ =  .676*
* N=     6       S.D.=      .50 PROB=  .005*
* PREDICTOR   3 LOCAL
* CODES  0
                     *
                     *
*GROUP  13       MEAN=      .23
* N=    22       S.D.=      .42
        LEVEL 2
                     *
                     *
*GROUP  19*FINAL MEAN=      .13 RSQ =  .676*
* N=    16       S.D.=      .33 PROB=  .008*
* PREDICTOR   3 LOCAL
* CODES  1
```

Exhibit 16 -- HIT/MISS TABLE--4/5 POLICY WITH CREDIT RATING

TABLE 2-CUMULATIVE FREQUENCY COUNTS

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
1.00	54	54	54	0	54	59	37	96	91	91
.99	5	5	59	0	59	54	37	91	96	96
.98	2	2	61	0	61	52	37	89	98	98
.97	1	1	62	0	62	51	37	88	99	99
.96	0	0	62	0	62	51	37	88	99	99
.95	1	1	63	0	63	50	37	87	100	100
.94	4	4	67	0	67	46	37	83	104	104
.93	7	7	74	0	74	39	37	76	111	111
.92	0	0	74	0	74	39	37	76	111	111
.91	0	0	74	0	74	39	37	76	111	111
.90	2	2	76	0	76	37	37	74	113	113
.89	0	0	76	0	76	37	37	74	113	113
.88	1	1	77	0	77	36	37	73	114	114
.87	0	0	77	0	77	36	37	73	114	114
.86	0	0	77	0	77	36	37	73	114	114
.85	0	0	77	0	77	36	37	73	114	114
.84	0	0	77	0	77	36	37	73	114	114
.83	2	2	79	0	79	34	37	71	116	116
.82	0	0	79	0	79	34	37	71	116	116
.81	0	0	79	0	79	34	37	71	116	116
.80	3	3	82	0	82	31	37	68	119	119
.79	0	0	82	0	82	31	37	68	119	119
.78	1	1	83	0	83	30	37	67	120	120
.77	3	3	86	0	86	27	37	64	123	123
.76	0	0	86	0	86	27	37	64	123	123
.75	0	0	86	0	86	27	37	64	123	123
.74	1	0	86	1	87	27	36	63	122	122
.73	0	0	86	1	87	27	36	63	122	122
.72	1	0	86	2	88	27	35	62	121	121
.71	1	1	87	2	89	26	35	61	122	122
.70	0	0	87	2	89	26	35	61	122	122
.69	1	1	88	2	90	25	35	60	123	123
.68	0	0	88	2	90	25	35	60	123	123
.67	1	0	88	3	91	25	34	59	122	122
.66	0	0	88	3	91	25	34	59	122	122
.65	0	0	88	3	91	25	34	59	122	122
.64	1	1	89	3	92	24	34	58	123	123
.63	1	1	90	3	93	23	34	57	124	124
.62	1	1	91	3	94	22	34	56	125	125
.61	0	0	91	3	94	22	34	56	125	125
.60	0	0	91	3	94	22	34	56	125	125
.59	3	3	94	3	97	19	34	53	128	128
.58	0	0	94	3	97	19	34	53	128	128
.57	0	0	94	3	97	19	34	53	128	128
.56	1	1	95	3	98	18	34	52	129	129
.55	2	2	97	3	100	16	34	50	131	131
.54	1	1	98	3	101	15	34	49	132	132
.53	0	0	98	3	101	15	34	49	132	132
.52	0	0	98	3	101	15	34	49	132	132
.51	0	0	98	3	101	15	34	49	132	132
.50	1	1	99	3	102	14	34	48	133	133

PREDICTED SCORE CUT-OFF	NUMBER NUM IN INTERVAL	NUMBER ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
.49	0	0	99	3	102	14	34	48	133	133
.48	0	0	99	3	102	14	34	48	133	133
.47	0	0	99	3	102	14	34	48	133	133
.46	0	0	99	3	102	14	34	48	133	133
.45	1	0	99	4	103	14	33	47	132	132
.44	1	1	100	4	104	13	33	46	133	133
.43	2	2	102	4	106	11	33	44	135	135
.42	1	1	103	4	107	10	33	43	136	136
.41	1	0	103	5	108	10	32	42	135	135
.40	0	0	103	5	108	10	32	42	135	135
.39	1	1	104	5	109	9	32	41	136	136
.38	0	0	104	5	109	9	32	41	136	136
.37	0	0	104	5	109	9	32	41	136	136
.36	0	0	104	5	109	9	32	41	136	136
.35	1	1	105	5	110	8	32	40	137	137
.34	2	1	106	6	112	7	31	38	137	137
.33	1	1	107	6	113	6	31	37	138	138
.32	0	0	107	6	113	6	31	37	138	138
.31	0	0	107	6	113	6	31	37	138	138
.30	0	0	107	6	113	6	31	37	138	138
.29	0	0	107	6	113	6	31	37	138	138
.28	0	0	107	6	113	6	31	37	138	138
.27	1	0	107	7	114	6	30	36	137	137
.26	1	0	107	8	115	6	29	35	136	136
.25	0	0	107	8	115	6	29	35	136	136
.24	1	0	107	9	116	6	28	34	135	135
.23	0	0	107	9	116	6	28	34	135	135
.22	0	0	107	9	116	6	28	34	135	135
.21	2	1	108	10	118	5	27	32	135	135
.20	0	0	108	10	118	5	27	32	135	135
.19	1	1	109	10	119	4	27	31	136	136
.18	2	0	109	12	121	4	25	29	134	134
.17	1	0	109	13	122	4	24	28	133	133
.16	1	1	110	13	123	3	24	27	134	134
.15	1	0	110	14	124	3	23	26	133	133
.14	2	0	110	16	126	3	21	24	131	131
.13	1	0	110	17	127	3	20	23	130	130
.12	4	0	110	21	131	3	16	19	126	126
.11	1	1	111	21	132	2	16	18	127	127
.10	1	0	111	22	133	2	15	17	126	126
.09	0	0	111	22	133	2	15	17	126	126
.08	1	0	111	23	134	2	14	16	125	125
.07	2	0	111	25	136	2	12	14	123	123
.06	2	1	112	26	138	1	11	12	123	123
.05	0	0	112	26	138	1	11	12	123	123
.04	1	0	112	27	139	1	10	11	122	122
.03	1	1	113	27	140	0	10	10	123	123
.02	0	0	113	27	140	0	10	10	123	123
.01	0	0	113	27	140	0	10	10	123	123
-.00	10	0	113	37	150	0	0	0	113	113

Exhibit 17 -- HIT/MISS TABLE--COMPOSITE POLICY WITH CREDIT RATING

TABLE 2--CUMULATIVE FREQUENCY COUNTS

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
1.00	28	28	28	0	28	85	37	122	65	65
.99	2	2	30	0	30	83	37	120	67	67
.98	2	2	32	0	32	81	37	118	69	69
.97	4	4	36	0	36	77	37	114	73	73
.96	3	3	39	0	39	74	37	111	76	76
.95	3	3	42	0	42	71	37	108	79	79
.94	2	2	44	0	44	69	37	106	81	81
.93	2	2	46	0	46	67	37	104	83	83
.92	10	10	56	0	56	57	37	94	93	93
.91	1	1	57	0	57	56	37	93	94	94
.90	3	3	60	0	60	53	37	90	97	97
.89	4	4	64	0	64	49	37	86	101	101
.88	0	0	64	0	64	49	37	86	101	101
.87	1	1	65	0	65	48	37	85	102	102
.86	3	3	68	0	68	45	37	82	105	105
.85	5	5	73	0	73	40	37	77	110	110
.84	2	2	75	0	75	38	37	75	112	112
.83	1	1	76	0	76	37	37	74	113	113
.82	1	1	77	0	77	36	37	73	114	114
.81	1	1	78	0	78	35	37	72	115	115
.80	0	0	78	0	78	35	37	72	115	115
.79	0	0	78	0	78	35	37	72	115	115
.78	1	1	79	0	79	34	37	71	116	116
.77	0	0	79	0	79	34	37	71	116	116
.76	0	0	79	0	79	34	37	71	116	116
.75	0	0	79	0	79	34	37	71	116	116
.74	0	0	79	0	79	34	37	71	116	116
.73	2	2	81	0	81	32	37	69	118	118
.72	2	2	83	0	83	30	37	67	120	120
.71	3	3	86	0	86	27	37	64	123	123
.70	1	0	86	1	87	27	36	63	122	122
.69	0	0	86	1	87	27	36	63	122	122
.68	0	0	86	1	87	27	36	63	122	122
.67	1	0	86	2	88	27	35	62	121	121
.66	0	0	86	2	88	27	35	62	121	121
.65	0	0	86	2	88	27	35	62	121	121
.64	1	1	87	2	89	26	35	61	122	122
.63	0	0	87	2	89	26	35	61	122	122
.62	1	1	88	2	90	25	35	60	123	123
.61	0	0	88	2	90	25	35	60	123	123
.60	2	1	89	3	92	24	34	58	123	123
.59	1	1	90	3	93	23	34	57	124	124
.58	1	1	91	3	94	22	34	56	125	125
.57	0	0	91	3	94	22	34	56	125	125
.56	0	0	91	3	94	22	34	56	125	125
.55	1	1	92	3	95	21	34	55	126	126
.54	0	0	92	3	95	21	34	55	126	126
.53	1	1	93	3	96	20	34	54	127	127
.52	1	1	94	3	97	19	34	53	128	128
.51	1	1	95	3	98	18	34	52	129	129
.50	1	1	96	3	99	17	34	51	130	130

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
.49	6	0	96	3	99	17	34	51	130	130
.48	0	0	96	3	99	17	34	51	130	130
.47	1	1	97	3	100	16	34	50	131	131
.46	1	1	98	3	101	15	34	49	132	132
.45	1	1	99	3	102	14	34	48	133	133
.44	0	0	99	3	102	14	34	48	133	133
.43	0	0	99	3	102	14	34	48	133	133
.42	1	0	99	4	103	14	33	47	132	132
.41	0	0	99	4	103	14	33	47	132	132
.40	0	0	99	4	103	14	33	47	132	132
.39	1	0	99	5	104	14	32	46	131	131
.38	0	0	99	5	104	14	32	46	131	131
.37	1	0	99	6	105	14	31	45	130	130
.36	0	0	99	6	105	14	31	45	130	130
.35	0	0	99	6	105	14	31	45	130	130
.34	1	1	100	6	106	13	31	44	131	131
.33	2	2	102	6	108	11	31	42	133	133
.32	0	0	102	6	108	11	31	42	133	133
.31	2	2	104	6	110	9	31	40	135	135
.30	0	0	104	6	110	9	31	40	135	135
.29	0	0	104	6	110	9	31	40	135	135
.28	2	2	106	6	112	7	31	38	137	137
.27	1	1	107	6	113	6	31	37	138	138
.26	1	0	107	7	114	6	30	36	137	137
.25	0	0	107	7	114	6	30	36	137	137
.24	0	0	107	7	114	6	30	36	137	137
.23	1	0	107	8	115	6	29	35	136	136
.22	0	0	107	8	115	6	29	35	136	136
.21	0	0	107	8	115	6	29	35	136	136
.20	1	0	107	9	116	6	28	34	135	135
.19	1	1	108	9	117	5	28	33	136	136
.18	0	0	108	9	117	5	28	33	136	136
.17	0	0	108	9	117	5	28	33	136	136
.16	0	0	108	9	117	5	28	33	136	136
.15	0	0	108	9	117	5	28	33	136	136
.14	2	1	109	10	119	4	27	31	136	136
.13	0	0	109	10	119	4	27	31	136	136
.12	3	0	109	13	122	4	24	28	133	133
.11	0	0	109	13	122	4	24	28	133	133
.10	3	2	111	14	125	2	23	25	134	134
.09	5	1	112	18	130	1	19	20	131	131
.08	0	0	112	18	130	1	19	20	131	131
.07	3	0	112	21	133	1	16	17	128	128
.06	3	0	112	24	136	1	13	14	125	125
.05	1	0	112	25	137	1	12	13	124	124
.04	0	0	112	25	137	1	12	13	124	124
.03	0	0	112	25	137	1	12	13	124	124
.02	1	1	113	25	138	0	12	12	125	125
.01	1	0	113	26	139	0	11	11	124	124
-.00	11	0	113	37	150	0	0	0	113	113

Exhibit 18 -- HIT/MISS TABLE--TENTATIVE APPROVAL WITHOUT CREDIT RATING

TABLE 2-CUMULATIVE FREQUENCY COUNTS

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
1.00	36	35	35	1	36	78	36	114	71	71
.99	5	5	40	1	41	73	36	109	76	76
.98	2	2	42	1	43	71	36	107	78	78
.97	1	1	43	1	44	70	36	106	79	79
.96	1	1	44	1	45	69	36	105	80	80
.95	2	1	45	2	47	68	35	103	80	80
.94	1	1	46	2	48	67	35	102	81	81
.93	3	3	49	2	51	66	35	99	84	84
.92	0	0	49	2	51	64	35	99	84	84
.91	3	3	52	2	54	61	35	96	87	87
.90	1	1	53	2	55	60	35	95	88	88
.89	3	3	56	2	58	57	35	92	91	91
.88	5	4	60	3	63	53	34	87	94	94
.87	3	3	63	3	66	50	34	84	97	97
.86	3	3	66	3	69	47	34	81	100	100
.85	4	4	70	3	73	43	34	77	104	104
.84	1	0	70	4	74	43	33	76	103	103
.83	1	1	71	4	75	42	33	75	104	104
.82	1	1	72	4	76	41	33	74	105	105
.81	2	1	73	5	78	40	32	72	105	105
.80	6	6	79	5	84	34	32	66	111	111
.79	3	1	80	7	87	33	30	63	110	110
.78	5	4	84	8	92	29	28	58	113	113
.77	2	0	84	10	94	29	27	56	111	111
.76	2	1	85	11	96	28	26	54	111	111
.75	2	2	87	11	98	26	26	52	113	113
.74	2	2	89	11	100	24	26	50	115	115
.73	1	0	89	12	101	24	25	49	114	114
.72	1	0	89	13	102	24	24	48	113	113
.71	2	2	91	13	104	22	24	46	115	115
.70	0	0	91	13	104	22	24	46	115	115
.69	2	1	92	14	106	21	23	44	115	115
.68	0	0	92	14	106	21	23	44	115	115
.67	0	0	92	14	106	21	23	44	115	115
.66	1	1	93	14	107	20	23	43	116	116
.65	2	2	95	14	109	18	23	41	118	118
.64	0	0	95	14	109	18	23	41	118	118
.63	0	0	95	14	109	18	23	41	118	118
.62	1	1	96	14	110	17	23	40	119	119
.61	2	2	98	14	112	15	23	38	121	121
.60	0	0	98	14	112	15	23	38	121	121
.59	1	1	99	14	113	14	23	37	122	122
.58	1	1	100	14	114	13	23	36	123	123
.57	0	0	100	14	114	13	23	36	123	123
.56	0	0	100	14	114	13	23	36	123	123
.55	2	1	101	15	116	12	22	34	123	123
.54	2	1	102	16	118	11	21	32	123	123
.53	0	0	102	16	118	11	21	32	123	123
.52	3	2	104	17	121	9	20	29	124	124
.51	0	0	104	17	121	9	20	29	124	124
.50	2	1	105	18	123	8	19	27	124	124

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
.49	0	0	105	18	123	8	19	27	124	124
.48	1	0	105	19	124	8	18	26	123	123
.47	0	0	105	19	124	8	18	26	123	123
.46	0	0	105	19	124	8	18	26	123	123
.45	0	0	105	19	124	8	18	26	123	123
.44	1	0	105	20	125	8	17	25	122	122
.43	1	0	105	21	126	8	16	24	121	121
.42	0	0	105	21	126	8	16	24	121	121
.41	0	0	105	21	126	8	16	24	121	121
.40	0	0	105	21	126	8	16	24	121	121
.39	0	0	105	21	126	8	16	24	121	121
.38	0	0	105	21	126	8	16	24	121	121
.37	0	0	105	21	126	8	16	24	121	121
.36	0	0	105	21	126	8	16	24	121	121
.35	0	0	105	21	126	8	16	24	121	121
.34	1	1	106	21	127	7	16	23	122	122
.33	0	0	106	21	127	7	16	23	122	122
.32	1	0	106	22	128	7	15	22	121	121
.31	1	1	107	22	129	6	15	21	122	122
.30	0	0	107	22	129	6	15	21	122	122
.29	0	0	107	22	129	6	15	21	122	122
.28	1	0	107	23	130	6	14	20	121	121
.27	1	1	108	23	131	5	14	19	122	122
.26	0	0	108	23	131	5	14	19	122	122
.25	1	1	109	23	132	4	14	18	123	123
.24	0	0	109	23	132	4	14	18	123	123
.23	0	0	109	23	132	4	14	18	123	123
.22	1	0	109	24	133	4	13	17	122	122
.21	0	0	109	24	133	4	13	17	122	122
.20	2	1	110	25	135	3	12	15	122	122
.19	0	0	110	25	135	3	12	15	122	122
.18	0	0	110	25	135	3	12	15	122	122
.17	1	0	110	26	136	3	11	14	121	121
.16	1	0	110	27	137	3	10	13	120	120
.15	0	0	110	27	137	3	10	13	120	120
.14	0	0	110	27	137	3	10	13	120	120
.13	1	1	111	27	138	2	10	12	121	121
.12	0	0	111	27	138	2	10	12	121	121
.11	0	0	111	27	138	2	10	12	121	121
.10	0	0	111	27	138	2	10	12	121	121
.09	0	0	111	27	138	2	10	12	121	121
.08	1	0	111	28	139	2	9	11	120	120
.07	0	0	111	28	139	2	9	11	120	120
.06	1	0	111	29	140	2	8	10	119	119
.05	0	0	111	29	140	2	8	10	119	119
.04	0	0	111	29	140	2	8	10	119	119
.03	0	0	111	29	140	2	8	10	119	119
.02	0	0	111	29	140	2	8	10	119	119
.01	0	0	111	29	140	2	8	10	119	119
.00	10	2	113	37	150	0	0	0	113	113

Exhibit 19 -- HIT/MISS TABLE--4/5 POLICY (LOCAL MODEL) WITH CREDIT RATING

TABLE 2-CUMULATIVE FREQUENCY COUNTS

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
1.00	5	4	4	1	5	109	36	145	40	40
.99	0	0	4	1	5	109	36	145	40	40
.98	74	74	78	1	79	35	36	71	114	114
.97	6	6	84	1	85	29	36	65	120	120
.96	0	0	84	1	85	29	36	65	120	120
.95	0	0	84	1	85	29	36	65	120	120
.94	3	3	87	1	88	26	36	62	123	123
.93	0	0	87	1	88	26	36	62	123	123
.92	0	0	87	1	88	26	36	62	123	123
.91	0	0	87	1	88	26	36	62	123	123
.90	0	0	87	1	88	26	36	62	123	123
.89	0	0	87	1	88	26	36	62	123	123
.88	3	3	90	1	91	23	36	59	126	126
.87	0	0	90	1	91	23	36	59	126	126
.86	0	0	90	1	91	23	36	59	126	126
.85	0	0	90	1	91	23	36	59	126	126
.84	0	0	90	1	91	23	36	59	126	126
.83	0	0	90	1	91	23	36	59	126	126
.82	2	2	92	1	93	21	36	57	128	128
.81	0	0	92	1	93	21	36	57	128	128
.80	0	0	92	1	93	21	36	57	128	128
.79	0	0	92	1	93	21	36	57	128	128
.78	0	0	92	1	93	21	36	57	128	128
.77	0	0	92	1	93	21	36	57	128	128
.76	5	4	96	2	98	17	35	52	131	131
.75	0	0	96	2	98	17	35	52	131	131
.74	0	0	96	2	98	17	35	52	131	131
.73	0	0	96	2	98	17	35	52	131	131
.72	0	0	96	2	98	17	35	52	131	131
.71	0	0	96	2	98	17	35	52	131	131
.69	0	0	96	2	98	17	35	52	131	131
.68	1	1	97	2	99	16	35	51	132	132
.67	0	0	97	2	99	16	35	51	132	132
.66	0	0	97	2	99	16	35	51	132	132
.65	0	0	97	2	99	16	35	51	132	132
.64	0	0	97	2	99	16	35	51	132	132
.63	0	0	97	2	99	16	35	51	132	132
.62	0	0	97	2	99	16	35	51	132	132
.61	0	0	97	2	99	16	35	51	132	132
.60	0	0	97	2	99	16	35	51	132	132
.59	0	0	97	2	99	16	35	51	132	132
.58	0	0	97	2	99	16	35	51	132	132
.57	0	0	97	2	99	16	35	51	132	132
.56	0	0	97	2	99	16	35	51	132	132
.55	0	0	97	2	99	16	35	51	132	132
.54	0	0	97	2	99	16	35	51	132	132
.53	0	0	97	2	99	16	35	51	132	132
.52	0	0	97	2	99	16	35	51	132	132
.51	0	0	97	2	99	16	35	51	132	132
.50	0	0	97	2	99	16	35	51	132	132

PREDICTED SCORE CUT-OFF	NONCUM NUM IN INTERVAL	NONCUM ACT 1 IN INTERVAL	NUMBER PRED 1 ACT 1	NUMBER PRED 1 ACT 0	NUMBER PRED 1	NUMBER PRED 0 ACT 1	NUMBER PRED 0 ACT 0	NUMBER PRED 0	NUMBER CORRECT	OBJECTIVE FUNCTION
.49	0	0	97	2	99	16	35	51	132	132
.48	0	0	97	2	99	16	35	51	132	132
.47	1	1	98	2	100	15	35	50	133	133
.46	0	0	98	2	100	15	35	50	133	133
.45	0	0	98	2	100	15	35	50	133	133
.44	0	0	98	2	100	15	35	50	133	133
.43	0	0	98	2	100	15	35	50	133	133
.42	0	0	98	2	100	15	35	50	133	133
.41	0	0	98	2	100	15	35	50	133	133
.40	0	0	98	2	100	15	35	50	133	133
.39	0	0	98	2	100	15	35	50	133	133
.38	0	0	98	2	100	15	35	50	133	133
.37	0	0	98	2	100	15	35	50	133	133
.36	0	0	98	2	100	15	35	50	133	133
.35	0	0	98	2	100	15	35	50	133	133
.34	1	1	99	2	101	14	35	49	134	134
.33	0	0	99	2	101	14	35	49	134	134
.32	0	0	99	2	101	14	35	49	134	134
.31	1	1	100	2	102	13	35	48	135	135
.30	0	0	100	2	102	13	35	48	135	135
.29	0	0	100	2	102	13	35	48	135	135
.28	0	0	100	2	102	13	35	48	135	135
.27	0	0	100	2	102	13	35	48	135	135
.26	0	0	100	2	102	13	35	48	135	135
.25	0	0	100	2	102	13	35	48	135	135
.24	0	0	100	2	102	13	35	48	135	135
.23	0	0	100	2	102	13	35	48	135	135
.22	0	0	100	2	102	13	35	48	135	135
.21	0	0	100	2	102	13	35	48	135	135
.20	0	0	100	2	102	13	35	48	135	135
.19	0	0	100	2	102	13	35	48	135	135
.18	0	0	100	2	102	13	35	48	135	135
.17	0	0	100	2	102	13	35	48	135	135
.16	0	0	100	2	102	13	35	48	135	135
.15	0	0	100	2	102	13	35	48	135	135
.14	0	0	100	2	102	13	35	48	135	135
.13	0	0	100	2	102	13	35	48	135	135
.12	0	0	100	2	102	13	35	48	135	135
.11	0	0	100	2	102	13	35	48	135	135
.10	0	0	100	2	102	13	35	48	135	135
.09	0	0	100	2	102	13	35	48	135	135
.08	0	0	100	2	102	13	35	48	135	135
.07	0	0	100	2	102	13	35	48	135	135
.06	0	0	100	2	102	13	35	48	135	135
.05	3	2	102	3	105	11	34	45	136	136
.04	17	9	111	11	122	2	26	28	137	137
.03	0	0	111	11	122	2	26	28	137	137
.02	0	0	111	11	122	2	26	28	137	137
.01	0	0	111	11	122	2	26	26	137	137
-.00	28	2	113	37	150	0	0	0	113	113

APPENDIX C

Users Description and Listing for AID4UT/AIDTRE and Ancillary Programs

Introduction:

AID4UT/AIDTRE are a pair of computer programs which implement the Automatic Interaction Detection (AID) Algorithm on the University of Texas CDC-6600 computer. They are adaptations of the AID-4 computer program developed by the Personnel Research Division of the Air Force Human Resources Laboratory, (AFHRL) Lackland AFB, Texas.

The basic AID algorithm was developed in 1964 by John A. Sonquist and James N. Morgan at the University of Michigan. They developed and exported the AID-2 program to various universities (including UT). The AID-4 version of the program from which the AID4UT/AIDTRE programs were developed represents an expansion of the AID-2 program by the AFHRL staff to include the following optional capabilities:

1. An option to select randomly a sample from the original group of observations and to process only that sample.

2. An option to select randomly a sample A and a sample B from the original group, to process sample A first and maintain a history of the splits, and then to force sample B to make the same splits as taken by sample A; i.e., single cross-validation by forced splitting.

3. An option to do the same as indicated in item 2, but in addition, to process subsequently sample B freely, and then to force Sample A to make the same splits as taken by sample B; i.e., double cross-validation by forced splitting.

4. An option to allow sample A and sample B to be provided by the user; i.e., no random sampling with AID-4, and to provide for single cross-validation. Actually double cross-validation can be accomplished by two separate runs and the switching of samples A and B.

5. An option to exclude cases with out-of range predictor values.

6. An option to specify the number of iterations to be printed in detail.

7. An option to print only a brief summary of the splitting involved.

8. An option to print <u>at each iteration</u> each of the following:

a. A t value testing the significance of the difference of the means of the resultant groups of a split.

b. A R^2 value at each split indicating the percentage of variance explained up through the current iteration.

c. A F value testing the significance of the increase in R^2 from the previous iteration.

d. A F value for a one-way analysis of variance conside ing all final groups at this iteration.

9. Various format changes for control cards, reports, etc., including additional messages explaining the direction of the split.

10. An option to plot the resultant splits.

The AID-4 program was operable on an IBM 7040 computer system. It was documented for use by the AFHRL in an internal technical report.

Modifications to the AID-4 program required for its implementation on the U. T. CDC-6600 included:

a. Development of a Fortran IV version of the plot routines.

b. Conversion from the program linkage used on the IBM 7040 into two separate programs with a tape file interface.

Further modifications were made to improve the utility of the program and included:

a. Addition of an optional skeleton tree plot that includes only the first eight plot levels.

b. Addition of a routine to display BSS/TSS profiles and to calculate a coefficient of Potential Explanatory Power (PEP) as proposed by B. M. Finifter of Michigan State University.

Purpose:

The purpose of this appendix is to provide a User's description of the AID4UT/AIDTRE program and to provide information on the ancillary programs used with AID4UT/AIDTRE in the process of Policy Capturing.

With the express permission of the Commander, AFHRL, Brooks AFB, Texas, the description of the input to the AID-4 program has been taken directly from the AFHRL technical report entitled "Automatic Interaction Detection, AID-4", July, 1971.

Additional information has been added to reflect the modifications made in the AID4UT/AIDTRE version of the program.

THE AID4UT ALGORITHM

The basic steps of the AID4UT algorithm for the "normal splitting" of groups into subgroups; i.e., no forced splitting, can be summarized below. The only additional logic required for forced splitting is concerned with the generation of a history file of normal splits to be imposed upon a different set of observations.

Initialization. Steps 1-4.

1. Read in all control cards (title card, format card or cards, description card, predictor cards, and criterion card) which define the predictor variables, the criterion variable, the stop criteria, etc., and establish all of the options to be processed. If an end-of-job card is read, terminate the job.

2. Print a summary of the control cards and all the variable definitions under "Control Card Parameters."

3. Read in the original input file, according to the options specified, and generate the recode categories (classes, responses, codes) for each case (observation). Edit each case and print each case with an out-of-range predictor values are to be excluded, also print any such cases under the same report. All cases listed under this report are eliminated at this point.

4. Identify all the remaining cases (assuming that no forced splitting or random sampling or partial use of the original input file is called for) as belonging to group number one. Group number one is the current candidate group for splitting and, indeed at the start, the only one. Continue to Step 8.

Test for Termination of the Algorithm. Step 5.

5. If either the current number of final groups (includes the candidate groups and the groups that cannot be split) equals MAXGP (a maximum number read in from the description card), or the number of candidate groups equals 90, terminate the algorithm by going to Step 25.

Determine Which Group Should be Selected for Attempted Partitioning. Steps 6-8.

6. Considering all groups obtained so far, select one of

them, group i, such that

a. The total sum of squares of that group (TSS_i) is greater than or equal to Pl (percentage read in from description card) percentage of the original total sum of squares (TSS_T); i.e., group number one.

b. The number of cases in group i is not less than twice the value of NMIN(the minimum number of cases required to be in both resultant groups for a candidate group to be split, read in from description card).

c. The group i has not already been split into two other groups.

d. There has been no previous failure to split the group i.

e. The total sum of squares of group i is not less than the sum of squares for any other group that meets the above four criteria (a-d).

7. If there is no such group i, terminate the algorithm by going to Step 26.

8. The group i selected is the current candidate group i which will be used for an attempted splitting (partitioning). Identify it with its group number (i) and print

$$N_i,\ W_i,\ \Sigma Y_i,\ \Sigma Y_i^2,\ TSS_i,\ \text{and } \overline{Y}_i,\ \text{and S.D.}_i.$$

Partition Scan Over All Predictors. Steps 9-14.

9. Set j, the predictor index, equal to one, BSS_{jMAX}, the maximum between sum of squares over all predictors of group i, equal to zero, and skip to Step 11.

10. Increment j by 1. If j is greater than the total number of predictors, go to Step 23, for the partition scan is complete. Otherwise, continue to Step 11.

11. For each category k of predictor j of group i, compute

$$N_{ijk},\ W_{ijk},\ \Sigma Y_{ijk},\ \Sigma Y_{ijk}^2,\ Y_{ijk},\ S.D_{ijk}.$$

12. If there are at least two categories k such that $N_{ijk} \neq 0$, continue to Step 13. Otherwise, print an appropriate message and return to Step 10, for predictor j has only one category.

13. If predictor j has been defined as monotonic, skip Step 14 (do not sort the Step 11 statistics), and go to Step 15 instead.

14. Sort the statistics produced in Step 11 together with the category identifiers for predictor j, into ascending order of the means ($\bar{Y}_{ijk}$'s).

Partition Scan Over the k_j Categories of Predictor j. Steps 15-19.

15. Set p, the category index, equal to zero, BSS_{kMAX}, the maximum between sum of squares over all categories of predictor j, equal to zero, and skip to Step 17.

16. Increase p by 1. If p is as large as k_j, where k_j is the number of categories within predictor j, then print the statistics for category k_j and go to Step 20, as all possible feasible splits to be examined are completed.

17. If either

$$\sum_{k=0}^{p} N_{ijk} < NMIN \text{ or}$$

$$(N_i - \sum_{k=0}^{p} N_{ijk}) < NMIN,$$

go back to Step 16, as this split cannot be made. Otherwise, compute BSS_p, the between sum of squares for the attempted split of group i on predictor j between the categories (0, 1, ..., p) and the adjacent categories (p+1, p+2, ..., k_j), and print the statistics for category p according to the print options specified.

18. If the BSS_p computed in Step 17 is less than BSS_{kMAX}, the current maximum between sum of squares, return to Step 16. Otherwise, continue to Step 19.

19. The BSS_p computed in Step 17 is the largest encountered so far for predictor j. Set $BSS_{kMAX} = BSS_p$, store p, and return to Step 16.

Determine the Best Predictor. Steps 20-21.

20. If BSS_{kMAX}, the maximum between sum of squares over all categories of predictor j, is equal to or larger than BSS_{jMAX}, the maximum between sum of squares computed for all previously scanned predictors, continue to Step 21. Otherwise, return to Step 10.

21. BSS_{kMAX} for predictor j identifies the best predictor and category tested so far of group i. Set $BSS_{jMAX} = BSS_{kMAX}$ and return to Step 10.

Determine Whether to Make Split. Steps 23-24.

23. If the maximum between sum of squares, BSS_{jMAX}, obtained after the scan of each category of each predictor of candidate group i is equal to or greater than P2(percentage read in from description card) percentage of the original total sum of squares, TSS_T, continue to Step 24. If not, eliminate group i as having failed in a split attempt, compute residuals as the options specify, and return to Step 6.

24. Split group i into two new groups, group NCF+1 and group NCF+2, where NCF is the current number of candidate and final groups, according to the predictor and category identified in Step 21. Print all information concerning the split including BSS_{jMAX}, BSS_{jMAX}/TSS_T, and the t value testing the significant difference between the means of group NCF+1 and group NCF+2. In addition, print a "Current Summary" indicating at this point the total number of candidate groups (including group NCF+1 and group NCF+2) and groups that cannot be split, TSS_T, BSS_T, WSS_T, R^2, the total proportion of the variance explained, an F value to test the significance of the reduction of the error sum of squares due to the new split, and an F value for a one-way analysis of variance for all final groups so far. Add group NCF+1 and group NCF+2 to the number of candidate groups and list all candidate groups, according to print options specified, before returning to Step 5.

Termination of the Algorithm. Steps 25-27.

25. Either the current number of final groups equals MAXGP or the number of candidate groups equals 90. Print an appropriate message, terminate the algorithm, and go to Step 27.

26. There are no more groups eligible for attempted partitioning. Print an appropriate message and go to Step 27.

27. Compute any residuals, according to options specified and return to Step 1.

The Program:

The compiled AID4UT/AIDTRE programs currently exists on Permanent File 4409 at the UT Computational Center. The AID4UT program processes the data set and determines the binary splits that best explain the variability in the data. On option, it prints out the results for each iteration and prints or punches a residuals list corresponding to the final groups. On option, it produces tape files for use by the AIDTRE program in graphically displaying the skeleton and detailed AID-Trees. The AIDTRE program operates in tandem with the AID4UT program.

Input Data:

The input for the AID4UT/AIDTRE programs consists of system control cards, program control cards and source data (either on cards or on tape). An example of a deck setup is shown in Figure C-1.

In the itemized description of the input data, a symbol of "R", "O", or "*", is given by each field. These symbols mean:

R = field required
O = field optional
* = exception described in field description.

ITEMIZED INPUT DESCRIPTION

AID-4 TITLE CARD

Card Columns		Description
1	R	The card type, punched equal to 1, which identifies the title card and initiates each AID-4 run. A title card is required at the beginning of each run, and if not, a terminal error will occur. AID-4 runs can be "stacked," i.e., more than one run made within the same job, only under the following conditions: a. Each run in the stack does not require an "unaccessible file" such as a tape file that is not currently mounted on a tape drive or such as a tape file that cannot be correctly positioned with the rewind option. b. The runs do not contain conflicting options such as an IRUN equal to 3.
2-49	O	The title which is printed at the beginning of the output under "Control Card Parameters." Forty-eight alphanumeric characters are provided for the user to identify a particular run.
50	R	IRUN, the run option, which can be summarized as the following:

Value of IRUN	Meaning
0	Use every case in the original input file for a normal AID-4 run; i.e., no forced splitting, provide BSS/TSS profiles.
1	Select a random sample A from the original input file, and use only the cases that belong to A for a normal AID-4 run. Then force those cases that belong to B to make the same splits as taken by A; i.e., single cross-validation.

	3	Select a random sample A and a random sample B from the original input file, and use only the cases that belong to A for a normal AID-4 run. Then force those cases that belong to B to make the same splits as taken by A. Then use only the cases that belong to B for a normal AID-4 run, and force those cases that belong to A to make the same splits as taken by B; i.e., double cross-validation.
	4	Given sample A and sample B, use only the cases that belong to A for a normal AID run. Then force those cases that belong to B to make the same splits as taken by A. Note that double cross-validation can be accomplished by submitting a second job with A and B switched.

51-56 R* NCPERM, the number of cases in the original input file.

NCPERM is optional if IRUN equals zero and the input file is a tape file ending with an end-of-file mark; i.e., NCPERM can be left blank, in which case all records in the input file will be processed until an end-of-file condition is reached.

NCPERM is also optional if IRUN equals zero and the input file is a card file providing an "end-of-file card" is placed behind the last data card.

NCPERM is not applicable if IRUN equals four.

57-58 0 NREELS, the number of reels for NCPERM.

If NREELS equals blanks or zeros, NREELS is assumed to be one.

If IRUN = 4, NREELS for sample A and B must be equal.

59-64	R*	NSAMA, the number of cases to be randomly selected in sample A.
		NSAMA is optional if IRUN equals four; i.e., NSAMA can be left blank in which case all cases in the sample A will be processed until an end-of-file condition is reached.
		NSAMA is not applicable if IRUN equals zero.
65-70	R*	NSAMB, the number of cases to be randomly selected in sample B.
		NSAMB is optional if IRUN equals four; i.e., NSAMB can be left blank in which case all cases in the sample B will be processed until an end-of-file condition is reached.
		NSAMB is not applicable if IRUN equals zero or one.
71-78	R*	NUM, an eight digit starting random number to be used in the random selection process when IRUN equals one, two, or three. NUM should not be a multiple of 17 or 5882353.
		NUM is not applicable if IRUN equals zero or four.
79-80	R	IFMT, the number of variable format cards to be read in next. The range of IFMT is $1 \leq \text{IFMT} \leq 4$.

AIR-4 FORMAT CARD OR CARDS

Card Columns		Description
1-78	R	The variable format beginning with a left parenthesis and ending with a right parenthesis specifies the integer type (I) conversion of the numerical data for the input file or files. Only integer type (I) fields can be defined within the parentheses, with X conversion to skip characters and the slash (/) to indicate the beginning of new records. The maximum number of input variables to be read in (as specified by NV in the description card) is 83 (80 fields for predictors, one field for the criterion, one field for identification, and one field for weights). Each predictor field read in, however, need not be used. Therefore, for certain sequences of runs, it may be advantageous to use the same format for all runs and control predictor selection with the predictor cards only. The only permissible characters in the input files to be read by the integer (I) type format are Ø - 9, +, -, and blank. Otherwise, a FORTRAN terminal error will occur. The number of format cards expected by AID-4 depends on the value of IFMT in the title card. At least one format card is required for each run with a maximum of no more than four. For IRUN equal to zero through three, the variable format card (or cards) specifies the arrangement of the data in the original input file. For IRUN equal to four, the variable format card (or cards) specifies <u>both</u> the given sample A and the given sample B. Therefore, the arrangement of sample A and sample B must be identical.
79-80		Not used.

AID-4 DESCRIPTION CARD

Card Columns		Description
1	R	The card type, punched equal to 3, which identifies the description card. A description card is required for each AID-4 run.
2-6	R	P1, the percentage of the original total sum of squares that must be contained in any given group if that group is to become a candidate group for splitting. That is, the total sum of squares in any given group must be greater than or equal to P1 percent of the original total sum of squares for that group to become a candidate group. Five decimal places are assumed (read with an F5.5 format). For example, to specify one percent, the user would punch 01000 in columns 2-6. Note that the decimal point is not actually punched in the card. The range of P1 is $.00001 \leq P1 < 1$. P1 punched equal to 00001 essentially deactivates this stopping criterion. (Cannot be left blank.)
7-11	R	P2, the best split of a given candidate group must reduce the original within (unexplained, error) sum of squares by P2 percentage of the original total sum of squares if that group is to be split. That is, the largest between sum of squares obtained after the scanning of all predictors of a given candidate group must be equal to or greater than P2 percentage of the original total sum of squares for that candidate group to be split. Furthermore, once a candidate group has filled the P2 requirement, that group will not be considered for splitting again, even though the group meets the P requirement. The range of P2 is such that $P2 \leq P1$. Ordinarily P2 is set equal to P1. (Cannot be left blank.) Note that P1 and P2 define an upper bound on the number of possible splits.
12-16	R	MAXGP, the maximum number of final groups (includes candidate groups and groups that cannot be split), regardless of the values of P1 and P2, that can be obtained before the algorithm will terminate.

If forced splitting is called for, i.e., IRUN greater than one, the maximum value for MAXGP is 90.

Ordinarily MAXGP is set high enough so that the stopping criterion is determined by P1, P2, and NMIN.

17-21 R NMIN, the minimum number of cases required to be in both resultant groups for a candidate group to be split. NMIN must be ≥ 2 or a fatal error will result.

22-26 R* KSTOP, the number of iterations to be printed in detail, i.e., the details of the scanning of all predictors for the candidate groups. After KSTOP is exceeded, only summary information is printed.

If KSTOP is left blank, KSTOP will be set to zero and only summary information will be printed.

If KSTOP equals -1, only a "Split Summary" is printed. (Minimum output)

27-29 R* NV, the number of variables to be read in from the input file. This number includes the predictor variables, criterion variable, the identification variable (if applicable), and the weight variable (if applicable). NV is required, must be less than or equal to 83, and must agree with the format as specified in the format cards.

30-32 R* NR, the field number of the case identification variable as determined from the format card or cards.

If residuals are called for (IRESID greater than zero), NR must be specified (0 < NR ≤ 83). Otherwise, NR can be left blank.

33 0 KRW, the rewind option which determines whether or not to rewind the original input file (a tape file) before the processing for a given run begins.

If KRW is equal to a zero or a blank, the original input file is not to be rewound.

If KRW is equal to one, the file is to be rewound.

If the original input file is a card file (ICARD equal to one), KRW is not applicable.

34 R IOPT, the exclude option which determines whether or not cases with out-of-range predictor values, as defined by the predictor cards, are excluded.

If IOPT is equal to a blank or a zero, the cases with out-of-range predictor values are not excluded. Predictor values less than the specified minimum values are assigned recode category 00 and predictor values greater than the specified maximum values are assigned the largest recode categories.

If IOPT is equal to one, the cases with out-of-range predictor values are excluded. The excluded cases are printed in the "Reject Summary" if the summary is called for; i.e., KREJ equal to one (see KREJ description).

IOPT is extremely important in that IOPT determines the format of the predictor cards (see description of AID-4 predictor cards).

35 0 KREJ, the reject summary print option which determines whether or not the "Reject Summary" is printed. The summary lists all those out-of-range cases that are excluded because of invalid predictor values or invalid criterion values.

If KREJ is a zero or a blank, the summary is not printed.

If KREJ is one, the summary is printed.

36 0 IRESID, the residual summary option which determines whether or not residuals (the criterion values minus the final group mean) are computed.

If IRESID is a zero or a blank, no residuals are computed.

If IRESID is one, the residuals are printed in the "Residual and/or Reject Summary." This summary is printed each time a final group is obtained; i.e., a group that cannot be split any more. Approximately 200 residuals are printed per page.

If IRESID is two, the residuals are punched into cards. Also IRUN must be less than two.

If IRESID is three, the residuals are written on Tape 8. Also IRUN must be less than two. With this option Tape 8 can be rewound and printed at job completion to give a concise list of residuals.

37 0 ICARD, the card input option which determines whether or not the original input file is cards.

If ICARD is zero or blank, the original input file is presumed to have been prepositioned on Tape 25.

If ICARD is one, the file is cards. However, IRUN must be less than three.

38 0 ITREE, the plot option which determines whether or not the normal splitting phases (1 and 3) are plotted. The forced splitting phases (2 and 4) cannot be plotted. If ITREE ≥ one, Tape 10 and Tape 14 are generated for use by AIDTRE.

If ITREE is zero or blank, no tapes are generated it ITREE is zero or blank, no tapes are generated.

If ITREE is one, only the detailed AID-TREE will be generated.

If ITREE is two, detailed and skeleton AID-TREES will be generated.

39 0 NOGO, the control card checking option which determines whether or not the run terminates after the printing of all control cards.

If NOGO is a zero or blank normal processing is assumed.

If NOGO is one, the run will terminate after the printing of the control card summary.

NOGO runs do not require any files and can be stacked with other runs.

40 0 IESS, the error sum of squares option which determines whether or not the error sum of squares for forced splits is printed.

If IESS is a zero or a blank, the error sum of squares is not printed.

If IESS is a one, the error sum of squares is printed.

41-80 Not used.

AID-4 PREDICTOR CARDS

Card Column		Description
1	R	The card type, punched equal to 4, which identifies the predictor card. At least one predictor card is required for each predictor (specified categories can require more than one card).
2-19	R	The name of the predictor which is printed at the beginning of the output under "Control Card Parameters" and each time a split is made on the predictor named. Eighteen alphanumeric characters are provided for the name.
20-22	R	The field number of the predictor as specified by the format card or cards.
23	R	KBL1, the first predictor specification variable, which determines the type of the predictor. If KBL1 is zero, the predictor is free-floating; i.e., the categories within the predictor are not sorted and remain in original recode order when the scanning process begins. If a predictor is free-floating, the between sum of squares obtained by the scanning process will be maximized over all possible splits for that predictor. If a predictor is monotonic, the between sum of squares obtained will be maximized only over the monotonic splits for that predictor. This does not, however, imply that all predictors should be typed as free-floating, for certain predictors make little sense as such. The type is dependent upon the nature of the predictor.
24	R	KBL2, the second predictor specification variable, which determines how the predictor is to be defined. The values of the predictor (as read by the variable format) must be positive or negative integers with a range of -99999 to 999999. AID-4 recodes the values of the predictor into categories numbered 00, 01, ..., 39. The maximum recode value is 39; i.e., the total number of recode values for any one predictor cannot exceed 40. Two methods of defining recodes are provided.

If KBL2 is zero, the predictor is defined to have equal intervals; i.e., the recode categories are determined from a minimum value; a maximum value, and an interval length. See description of columns 25-80.

If KBL2 is one, the predictor is defined to have specified categories; i.e., the categories are determined from specified input and recode values. See description of columns 25-80.

25-80 The format of columns 25-80 is dependent upon KBL2 and the exclude option, IOPT (column 34 of the description card), and can be summarized as the following four possibilities:

I. KBL2 equal to zero and IOPT equal to a blank or zero; i.e., equal intervals and cases with out-of-range predictor values not excluded.

I 25-30 R The minimum value, which determines the predictor values to be included in the first category; i.e., category 00. Any predictor value less than or equal to the minimum value will be included in category 00. Also any predictor value less than one interval greater than the minimum value will be included in category 00.

I 31-36 R The maximum value, which determines the predictor values to be included in the last category. Any predictor value greater than or equal to the maximum value will be included in the last category.

I 37-42 R The interval length, which together with the minimum and maximum values, determines the number of recode categories. The number of categories is computed by

$$((\text{max-min})/\text{interval}) + 1,$$

which must be less than 40. Note that remainders after division are truncated.

For predictor values greater than the minimum value and less than the maximum value, the recode categories are computed by

$$\frac{\text{(predictor value - min)}}{\text{interval}}$$

Note that remainders are truncated.

I 43-45 R Specified recode category 1, the value of which is assigned to any predictor value equal to specified input value 1 (columns I 46-51). Provision is made for reassigning up to three values of the predictor (using specified recode categories 1, 2, and 3). Note, however, that the specified input values 1, 2, and 3 should be greater than or equal to the minimum value and less than or equal to the maximum value, as the range check for minimum and maximum values is performed first. If a predictor value is less than the minimum, the value is set equal to the minimum. If a predictor value is greater than the maximum, the value is set equal to the maximum.

If the recode is not used, a -1 must be punched in columns I 44-45.

I 46-51 R Specified input value 1.

I 52-54 R Specified recode category 2, the value of which is assigned to any predictor value equal to specified input value 2 (columns I 55-60).

If the recode is not used, a -1 must be punched in columns I 53-54.

I 55-60 R Specified input value 2.

I 61-63 R Specified recode category 3, the value of which is assigned to any predictor value equal to specified input value 3 (columns I 64-69).

If the recode value is not used, a -1 must be punched in columns I 62-63.

I 64-69 R Specified input value 3.

70-80 Not used.

II. KBL2 equal to zero and IOPT equal to one; i.e., equal intervals and cases with out-of-range predictor values excluded.

II 25-30 R The minimum value by which any predictor value less than the minimum value is excluded.

II 31-36 R The maximum value by which any predictor value equal to or greater than the maximum is excluded.

II 37-42 R The interval length, which together wit the minimum and maximum values, determines the number of recode categories. The number of categories is computed by

$$((\text{max-min})/\text{interval}),$$

which must be less than 40. Note that remainders after division are truncated.

For predictor values greater than or equal to the minimum value and less than the maximum values the recode categories are computed by

$$\frac{(\text{predictor value - min})}{\text{interval}}$$

Note that remainders are truncated.

II 43-45 R Specified recode category 1, the value of which is assigned to any predictor value equal to specified input value 1 (columns II 46-51). Provision is made for reassigning up to three values of the predictor (using specified recode categories 1, 2, and 3). Note, however, that the specified input values 1, 2, and 3 should be greater than or equal to the minimum value and less than the maximum value, as the range check for minimum and maximum values is performed first. If a predictor value is less than the minimum or greater than or equal to the maximum, the case is excluded.

If the recode is not used, a -1 must be punched in columns 44-45.

II 46-51	R	Specified input value 1. Note that this value cannot be greater than or equal to the maximum, or less than the minimum.
II 52-54	R	Specified recode category 2, the value of which is assigned to any predictor value equal to specified input value 2 (columns II 55-60). If the recode is not used, a -1 must be punched in columns II 53-54.
II 55-60	R	Specified input value 2. Note that the value cannot be greater than or equal to the maximum, or less than the minimum.
II 61-63	R	Specified recode category 3, the value of which is assigned to any predictor value equal to specified input value 3 (columns II 64-69). If the recode value is not used, a -1 must be punched in columns II 62-63.
II 64-69	R	Specified input value 3. Note that the value cannot be greater than or equal to the maximum, or less than the minimum.
II 70-80	R	Not used.

III. KBL2 equal to one and IOPT equal to a blank or zero, i.e., <u>specified categories and cases with out-of-range predictor values not excluded</u>.

III 25-27	R	Specified recode category 1, the value of which is assigned to any predictor value equal to or less than specified input value 1 (columns III 28-33).
III 28-33	R	Specified input value 1.
III 34-36	R	Specified recode category 2, the value of which is assigned to any predictor value greater than specified input value 1 and less than or equal to specified input value 2 (columns III 37-42).
III 37-42	R	Specified input value 2.

III	43-45	R	Specified recode category 3, the value of which is assigned to any predictor value greater than specified input value 2 and less than or equal to specified input value 3 (columns III 46-51).
III	46-51	R	Specified input value 3.
III	52-54	R	Specified recode category 4, the value of which is assigned to any predictor value greater than specified input value 3 and less than or equal to specified input value 4 (columns III 55-60).
III	55-60		Specified input value 4.
III	61-63	R	Specified recode category 5, the value of which is assigned to any predictor value greater than specified input value 4 and less than or equal to specified input value 5 (columns III 64-69).
III	64-69	R	Specified input value 5.
III	70-80		Not used.

If more than five categories are required, additional predictor cards can be used by duplicating columns III 1-24 and continuing in columns III 25-69 respectively. As many cards per predictor as necessary can be used as long as the maximum of 40 recode categories is not exceeded; i.e., 00-39.

The last specified input value must always equal 999999 and any predictor value greater than the previous specified input value is assigned the last specified recode category. For example, if specified input value 4 equaled 999999, any predictor value greater than input value 3 would be assigned recode category 4.

IV. KBL2 equal to one and IOPT equal to one; i.e., specified categories and cases with out-of-range predictor values excluded.

IV	25-27	R	Blank (except for continuation cards; see next to last paragraph below.)

IV	28-33	R	The lower bound by which any predictor value less than or equal to the lower bound is excluded.
IV	34-36	R	Specified recode category 2, the value of which is assigned to any predictor value greater than the lower bound and less than or equal to specified input value 2 (columns IV 37-42).
IV	37-42	R	Specified input value 2.
IV	43-45	R	Specified recode category 3, the value of which is assigned to any predictor value greater than specified input value 2 and less than or equal to specified input value 3 (columns IV 46-51).
IV	46-51	R	Specified input value 3.
IV	52-54	R	Specified recode category 4, the value of which is assigned to any predictor value greater than specified input value 3 and less than or equal to specified input value 4 (columns IV 55-60).
IV	55-60	R	Specified input value 4.
IV	61-63	R	Specified recode category 5, the value of which is assigned to any predictor value greater than specified input value 3 and less than or equal to specified input value 5 (columns IV 64-69).
IV	64-69	R	Specified input value 5.
IV	70-80		Not used.

If more than three categories are required, additional predictor cards can be used by duplicating columns IV 1-24 and continuing in columns IV 25-69 respectively. On second and following continuation cards, start next specified recode category in columns IV 25-27. <u>As many cards per predictor as necessary can be used as long as the maximum of 40 recode categories is not exceeded; i.e., 00-39</u>.

The last specified input value must always equal 999999 and the previous input value acts as the upper bound by which any predictor value greater than the upper bound is excluded. The last specified recode category (before the 999999's) must be left blank.

AID-4 CRITERION CARD

Card Columns		Description
1	R	The card type, punched equal to 5, which identifies the criterion card. One criterion card is required for each run.
2-19	R	The name of the criterion which is printed at the beginning of the output under "Control Card Parameters " Eighteen alphanumeric characters are provided for the name.
20-22	R	The field number of the criterion as specified by the format card or cards.
23-24	0	The field number of the weight as specified by the format card or cards. If this field is left blank, the weight is set to one.
25-30	R	YMAX, the maximum allowable value of the criterion variable. If the value of the criterion is greater than YMAX, the case is excluded. The maximum value that can be specified for YMAX is 999999.
31-36	R	MD1, the first deletion value, by which any criterion value equal to MD1 will be excluded. MD1 can be deactivated by setting MD1 greater than YMAX.
37-42	R	MD2, the second deletion value, by which any criterion value equal to MD2 will be excluded. MD2 can be deactivated by setting MD2 greater than YMAX.
49-80		Not used.

AID-4 END-OF-JOB CARD

Card Columns		Description
1	R	The card type, punched equal to 9, which identified the end-of-job card and terminates the job. The purpose of the end-of-job card is to indicate the end of a run executed within a given job. The card should be thought of strictly as a control card such as the title card, format card, etc., and not as part of the original input file.
2-80		Not used.

AID-4 LIMITATIONS

For the current version of AID-4, the following list of limitations apply:

1. The maximum number of input variables to be read in is 83 (80 fields for predictors, one field for the criterion, one field for identification, and one field for weights).

2. The maximum number of predictor variables is 80.

3. The maximum number of generated categories for all predictor variables taken together must not exceed 700. The total number of categories for all predictors, N_C, can be computed by the following formula:

$$N_C = 4N_E + 2N_{SR} + 2N_{SC}$$

where N_E is the number of predictor variables with equal intervals, N_{SR} is the number of specified recode categories for predictors with equal intervals, and N_{SC} is the number of specified recode categories for predictors not with equal categories.

4. The maximum number of criterion variables for any one analysis is one.

5. The maximum number of candidate groups at any given iteration is 90.

6. The maximum number of cards for the variable input format is four.

7. The maximum number of final groups, including the candidate groups and the groups that cannot be split, is 99999, the maximum value of MAXGP. Note that the number of final groups is also dependent upon P1, P2, and NMIN.

8. The range of the recode categories is 0 to 39.

9. The range of the predictor values is -99999 to 999999.

10. The range of the criterion values is -99999 to 999999.

11. The range of the weight values is determined by the variable format.

12. The minimum value for NMIN is 2.

AID-4 LIST OF TERMINAL ERRORS

Error Number	Description
1	AID-4 title card (card type 1) expected.
2	IRUN out-of-range (greater than 4).
3	IRUN equal to 1, 2, or 3 and NCPERM equal to zero.
4	IRUN equal to 1, 2, or 3 and NSAMA equal to zero.
5	IRUN equal to 2 or 3 and NSAMB equal to zero.
6	IRUN equal to 1, 2, or 3 and NUM equal to zero.
7	IRUN equal to 1, 2, or 3 and NUM is a multiple of 17.
8	IRUN equal to 1, 2, or 3 and NUM is a multiple of 5882353.
9	IFMT out-of-range (equal to zero or greater than 4).
10	AID-4 description card (card type 3) expected.
11	P1 out-of-range (greater than 1).
12	P2 out-of-range (greater than 1).
13	MAXGP out-of-range (less than 1).
14	KRW out-of-range (greater than 1).
15	IOPT out-of-range (greater than 1).
16	KREJ out-of range (greater than 1).
17	IRESID out-of-range (greater than 3).
18	IRUN greater than 1 and IRESID greater than 1.
19	The field number of a predictor card is greater than NV.
20	The field number of a predictor card is less than 1.
21-24	Either KBL1 or KBL2 out-of-range (not equal to zero or 1).
25	Interval length out-of-range (less than 0).
26	Number of categories over all predictors greater than 700.
27	Number of categories over all predictors greater than 699.
28	AID-4 predictor card (card type 4) expected.
29	Continuation card incorrect.
30	Number of categories greater than 40.
31	Total number of categories greater than 700.
32	AID-4 criterion card (card type 5) expected.
33	The field number of the criterion card is greater than NV.
34	The field number of the criterion card is less than 1.
35	The weight field in the criterion card is greater than NV.
36	The weight field in the criterion card is negative.

37	YMAX is less than or equal to YMIN.
38	NV is greater than 83.
39	Not used.
40	Interval length is less than zero.
41	ICARD greater than 1.
42	IRUN greater than 2 and ICARD equal to 1.

An Example Problem:

Figure C-1 shows the input deck for a problem with 12 data points. The first four cards are the system control cards required to execute AID4UT/AIDTRE. There are three predictor variables, (X_1, X_2, X_3). Figure C-2 shows the standard output set produced by all AID4UT runs. Figure C-3 shows the detailed output for the first iteration of the splitting process. This detailed output is optional and controled by the input parameter, KSTOP.

Figure C-4 shows the residuals table for groups 8 and 9 and the summary information available at the end of the splitting process. It also shows the BSS/TSST and BSS/TSS(i) profiles and the corresponding ranking of the predictor variables based on the PEP coefficient (RECSUM). This latter coefficient was proposed by Finifter and represents the average explanatory power of a predictor over all groups in the AID-Tree. This output is only available with IRUN $\leq$ 1.

Figure C-5 presents the skeleton version of the AID-Tree that is produced when ITREE = 2. Figure C-6 presents the detailed AID-Tree that is produced when ITREE $\leq$ 2. If an AID-Tree is not desired for a run, removal of the EXECPF(4409, AIDTRE) card will eliminate the tree but will not affect the output shown in Figures C-2 through C-4.

Setting of Parameters in AID4UT/AIDTRE:

In initial exploratory analysis runs on a data set, the following parameter settings are recommended:

IRUN = 0, P1 = .0001, P2 = .0001, MAXGP $\simeq$ 100, NMIN = 2,

KSTOP = 5, IRESID = 0, ITREE = 2.

These settings allow maximum splitting to occur and detection of any coding errors and refinement of predictor categories (if these categories were arbitrarily defined originally). These parameters do allow considerable overfitting to occur if the data set is small and further runs with restricted parameter values will generally be required.

EXAMPLE INPUT DECK

```
RFL,130000.
EXECPF,4408,AID4UT
RFL,150000.
EXECPF,4408,AIDTRE
7/8/9
```

SYSTEM CONTROL CARDS

```
1   AID4UT DEMONSTRATION RUN                    0    12                   1
(I2,3I1,I4)
3000010000l   12    2    4  5  1    112
4  X1                210    1    2    1 -1    1 -1    1 -1
4  X2                310    1    2    1 -1    1 -1    1 -1
4  X3                410    1    4    1 -1    1 -1    1 -1
5  RESP (Y)          5    500   -50   501   501
 1111 125
 2113 245
 3114 305
 4121 205
 5123 365
 6124 445
 7211   3
 8213  -1
 9214  -3
10221   9
11223   9
12224   9
9
```

SOURCE DATA

PROGRAM CONTROL CARDS

FIGURE C-1

```
UNIVERSITY OF TEXAS VERSION OF AID4 AS ADAPTED FROM AFHRL,LACKLAND AFB
MODIFIED FOR CDC6600 BY CAPT LL GOUCH, OCT 72

ONTROL CARD PARAMETERS


TITLE CARD

CARD
TYPE                    TITLE                                  IRUN  NCPERM  NREELS  NSAMA  NSAMB   NUM   IFMT
 1     AID4UT DEMONSTRATION RUN                                  8     12      0       0      0      0     1

FORMAT CARD OR CARDS


(I2,3I1,I4)

DESCRIPTION CARD

CARD
TYPE    P1       P2    MAXGP  NMIN  KSTOP  NV  NH  KRW  IOPT  KREJ  IRESID  ICARD  ITREE  NOGO  IESS  IPSQ1  IPSQ2
 3    .00001   .00001   12     2     4     5   1   0     0     0      1      1      2      0     0     0      0


   VARIABLE        FIELD NO.  RECODE      CORRESPONDS TO      TYPE


                                                             MONOTONIC
 1    A1                2        0    LESS THAN      2
                                 1         2 OR OVER

                                                             MONOTONIC
 2    A2                3        0    LESS THAN      2
                                 1         2 OR OVER

                                                             MONOTONIC
 3    A3                4        0    LESS THAN      2
                                 1         2 TO      2
                                 2         3 TO      3
                                 3         4 OR OVER

DEPENDENT VARIABLE-Y IN FIELD NO.   5 IS    RESP (Y)

   YMAX. IS    500...YMIN. IS   -50...EXCLUDE    501    501.    WEIGHT IN FIELD NO.   0

DO NOT SPLIT UNLESS TSS MORE THAN       2.5127500      AND BSS MORE THAN       2.5127500
             12 VALID CASES. GROUP 1 VALUES ARE-

  GROUP    N    TOTAL WEIGHT      SUM OF Y      SUM Y-SQUARE        T S S          MEAN          STD. DEV.
    1     12    12.000000        1950.0000      566150.00        251275.00      162.50000      144.70614
```

FIGURE C-2

```
** STEP NO.= 1        PARENT GROUP =  1 **
 TRY ON PREDICTOR  1      X1
   CODE    N      TOTAL WEIGHT        SUM OF Y        SUM Y-SQUARE          MEAN            STD. DEV.          B S S            W S S
    0      6      6.0000000          1690.0000        541950.00           281.66667         104.82790
   ****************************************************************************************************   170408.33         80866.667
    1      6      6.0000000          260.00000        26200.000           43.333333         49.888765

MAX. BSS=     170408.33         BSS/TSS =     .67817   BETWEEN CODES    0
                                                       AND CODES        1
 TRY ON PREDICTOR  2      X2
   CODE    N      TOTAL WEIGHT        SUM OF Y        SUM Y-SQUARE          MEAN            STD. DEV.          B S S            W S S
    0      6      6.0000000          665.00000        170575.00           110.83333         127.66352
   ****************************************************************************************************   32033.333         219241.67
    1      6      6.0000000          1285.0000        397575.00           214.16667         142.81155

MAX. BSS=     32033.333         BSS/TSS =     .12748   BETWEEN CODES    0
                                                       AND CODES        1
 TRY ON PREDICTOR  3      X3
   CODE    N      TOTAL WEIGHT        SUM OF Y        SUM Y-SQUARE          MEAN            STD. DEV.          B S S            W S S
    0      4      4.0000000          450.00000        66650.000           112.50000         63.294945
   ****************************************************************************************************   15000.000         236275.00
    2      4      4.0000000          690.00000        201450.00           172.50000         143.54617
   ****************************************************************************************************   9600.0000         241675.00
    3      4      4.0000000          810.00000        400050.00           202.50000         184.40784

MAX. BSS=     15000.000         BSS/TSS =     .05970   BETWEEN CODES    0
                                                       AND CODES        2  3


*****************************************************************************************************************************
SPLIT GROUP   1 ON PREDICTOR   1    X1                 INTO GROUP   2 WITH CODES  0
                                                       AND GROUP    3 WITH CODES  1

        BSS  =     170408.33       BSS/TSS  =    .67817   T-VALUE  6.59

*****************************************************************************************************************************

                                                    CURRENT SUMMARY


NCF      TOTAL TSS       TOTAL BSS       TOTAL WSS       R-SQUARED       R         F-RSQ    DF1   DF2      F-ANOVA   DF1    DF2

 2      251275.00       170408.33       80866.667       .67817464      .82351     21.0728    1     10      21.0728    1      10

*****************************************************************************************************************************

  CANDIDATE GROUPS ARE  AS FOLLOWS.

  GROUP    N      TOTAL WEIGHT        SUM OF Y        SUM Y-SQUARE          T S S            MEAN            STD. DEV.
    2      6      6.0000000          1690.0000        541950.00          65933.333         281.66667        104.82790
    3      6      6.0000000          260.00000        26200.000          14933.333         43.333333        49.888765
```

FIGURE C-3

RESIDUAL AND/OR REJECT SUMMARY

GROUP	ID	WT	RES	GROUP	ID	WT	RES	GROUP	ID	WT	RES	GROUP	ID	WT	RES
6	7	1	33	8	6	1	-7	8	8	1	-27	9	10	1	0
9	11	1	0	9	12	1	6								

FINAL SUMMARY

NCF	TOTAL TSS	TOTAL BSS	TOTAL WSS	R-SQUARED	R	F-ANOVA	DF1	DF2
5	251275.88	241288.33	10866.667	.95983765	.97976	41.9319	4	7

END OF AIDOUT - DATA FOR TREE ON TAPES 10 - 14

ANALYSIS WITH BSS/TSST

TRIAL/GRP		1 X1	2 X2	3 X3
1	1	.678175	.127483	.955486
2	2	0.000053	.876676	.162535
3	5	0.000000	.667257	.010581
4	3	0.000000	.052061	.063316
SUBSUM		.67817	.32342	.24502
TRIALS		1	4	4
MEANBS		.67817	.08085	.06125
S D		0.00000	.02832	.06196
CF VAR		0.00000	.35031	1.01146
VAR		0.00000	.00080	.00384
RECSUM		2.71276	.32342	.24502

ANALYSIS WITH BSS/TSS(1)

TRIAL/GRP		1 X1	2 X2	3 X3
1	1	.678175	.127483	.059600
2	2	0.500000	.292214	.619312
3	5	0.000059	.771089	.223746
4	3	0.000000	.875000	.055096
SUBSUM		.67817	2.06630	.95856
TRIALS		1	4	4
MEANBS		.67817	.51660	.23964
S D		0.00000	.31436	.22944
CF VAR		0.00000	.60852	.95746
VAR		0.00000	.09882	.05264
RECSUM		2.71270	2.06639	.95856

SUMMARY TABLE 3 VARIABLES, NO OF SUBGROUPS IS 4 GRAND MEAN = 1.093711 ROUND CRIT. = .010TSST =251275.0

		SUB(VARIANT) SUMBSS/TSST	RECONSTRUCTED SUMBSS/TSST	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	X1	.67817	2.71278	0.00028	0.00039	0.00000	248.02781	1	1.0
2	X2	.32342	.32342	.02832	.35031	.00089	29.57063	4	2.0
3	X3	.24502	.24502	.06196	1.01146	.00384	22.40234	4	3.0

SUMMARY TABLE 3 VARIABLES, NO OF SUBGROUPS IS 4 GRAND MEAN = 1.912547 ROUND CRIT. = .010TSST =251275.0

		SUB(VARIANT) SUMBSS/TSS(1)	RECONSTRUCTED SUMBSS/TSS(1)	S D	COEFF VAR	VARIANCE PCT	GRD MEAN	N GROUPS	RANK
1	X1	.67817	2.71278	0.00000	0.00000	0.00000	141.83095	1	1.0
2	X2	2.06639	2.06639	.31436	.60852	.09882	108.04372	4	2.0
3	X3	.95856	.95856	.22944	.95746	.05264	50.11932	4	3.0

FIGURE C-4

TREE STRUCTURE OF FIRST EIGHT LEVELS, GROUP NUMBER IS GIVEN IN CENTER OF THE BOX, THE PAIR MEMBER WITH THE HIGHER MEAN IS ON TOP

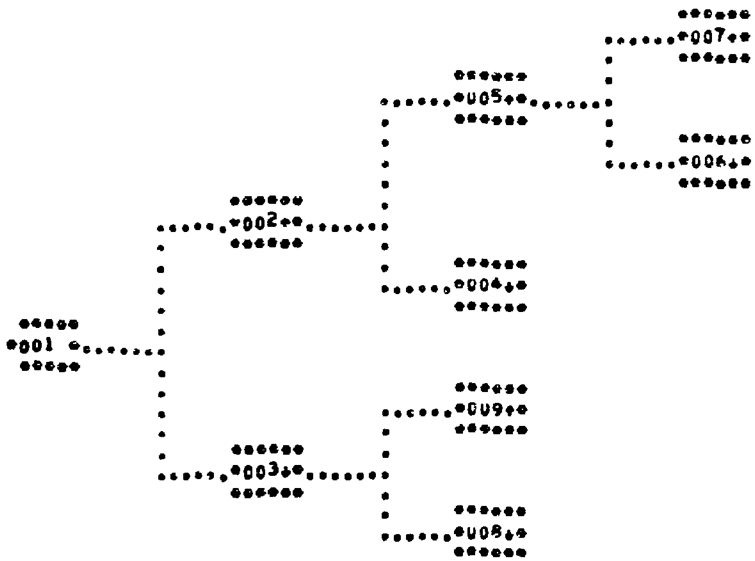


FIGURE C-5

```
GROUP   2        MEAN=   281.67 RSQ =  .678
 N=     6        S.D.=   104.83 PROB=  .001
 PREDICTOR  1    X1
 CODES  0

GROUP   5        MEAN=   340.00 RSQ =  .841
 N=     4        S.D.=    73.99 PROB=  .014
 PREDICTOR  3    X3
 CODES  2  3

GROUP   7*FINAL  MEAN=   405.00 RSQ =  .908
 N=     2        S.D.=    40.00 PROB=  .042
 PREDICTOR  2    X2
 CODES  1

GROUP   6*FINAL  MEAN=   275.00 RSQ =  .918
 N=     2        S.D.=    30.00 PROB=  .042
 PREDICTOR  2    X2
 CODES  0

GROUP   4*FINAL  MEAN=   165.00 RSQ =  .841
 N=     2        S.D.=    40.00 PROB=  .014
 PREDICTOR  3    X3
 CODES  0

GROUP   1        MEAN=   162.50
 N=    12        S.D.=   144.71
      LEVEL 1

GROUP   9*FINAL  MEAN=    90.00 RSQ =  .960
 N=     3        S.D.=     0.00 PROB=  .020
 PREDICTOR  2    X2
 CODES  1

GROUP   3        MEAN=    43.33 RSQ =  .678
 N=     6        S.D.=    49.89 PROB=  .001
 PREDICTOR  1    X1
 CODES  1

GROUP   8*FINAL  MEAN=    -3.33 RSQ =  .960
 N=     3        S.D.=    24.94 PROB=  .020
 PREDICTOR  2    X2
 CODES  0
```

FIGURE C-6

In detailed analysis runs, the critical parameters become P1, P2, MAXGP and NMIN. These values are dependent upon the size of the data set. In cases where there are less than 500 data units, the following values are suggested.

$P1 \geq .01 \qquad P2 \geq .005 \qquad MAXGP \leq 90$

NMIN $\geq$ 5 percent of data set, KSTOP ≤ 10

Other parameter settings are dependent only on the output desires of the analyst. For example, the residuals list may only be desired on final runs.

Use of AID4UT/AIDTRE with other Computer Programs as a Policy Capturing System:

In the process of Policy Capturing, various ancillary computer programs were used with AID4UT/AIDTRE to analyze and develop regression models. These programs were:

Program	Purpose	Documentation
GCORR	Contingency Table Generation for Data Analysis	Anderberg (1971)
REG REJ (EDSTAT-J)	Multiple regression model determination and vross validation	Jennings (1971)
MCA (SOCLIB)	Multiple Classification Analysis	UT-Computation Center
STEP-01	Stepwise Regression	CFHR, ESB, UT Austin
NCRIT	Prediction of Binary Vector	(Listing Included)

Figure C-7 is a Process Diagram for using the various programs in the Policy Capturing process. Alternative coefficient determination schemes allow the analyst considerable flexibility. The

PROCESS DIAGRAM

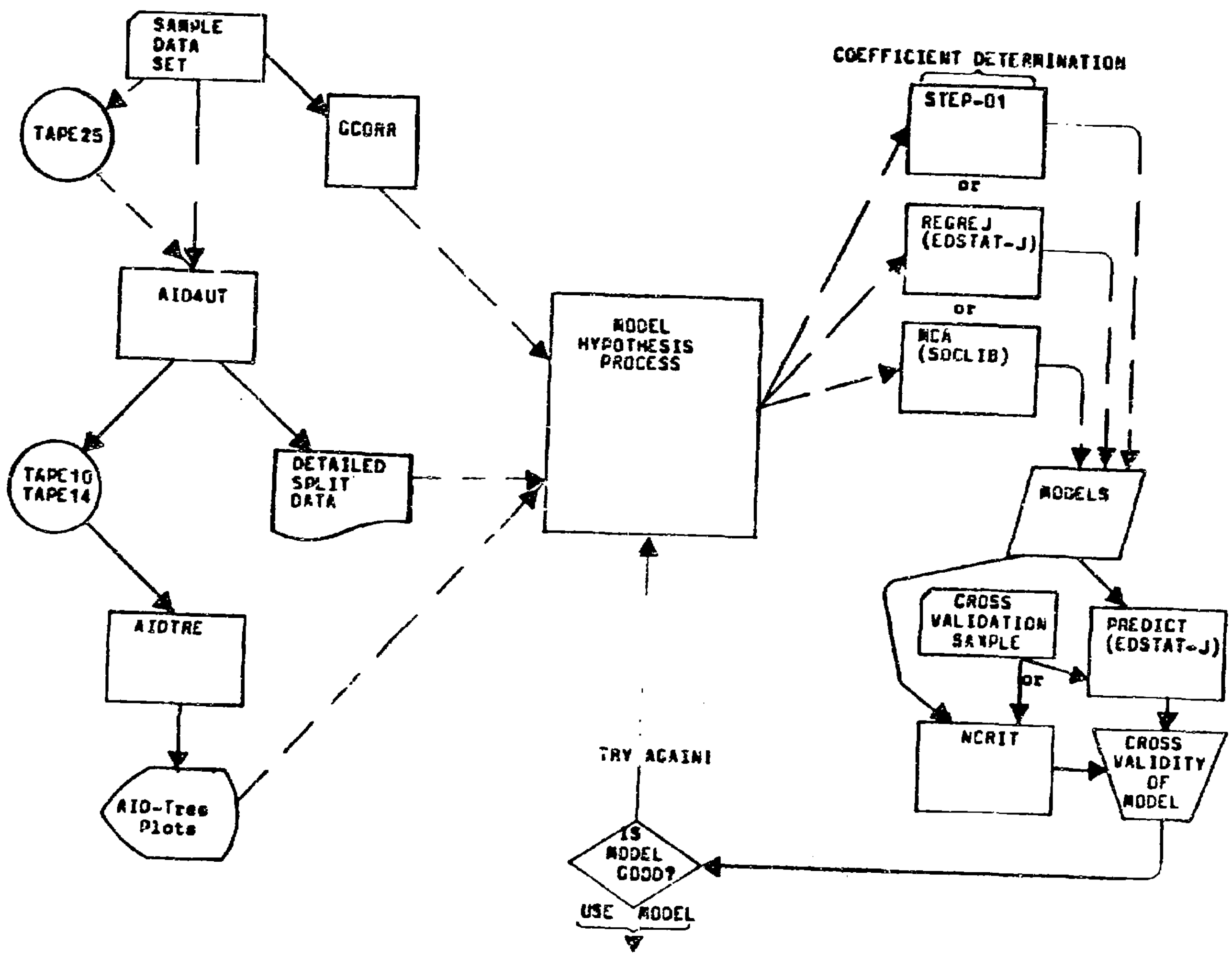


FIGURE C-7

use of the EDSTAT-J system represents a particularly important option since the analyst can write a driver program which uses the standard statistical subroutines to perform a statistical analysis that is tailored to his specific needs. As an example, the listing for a driver program named CRV is given later in this appendix. This is the routine that was written to perform the interative determination/cross-validation of regression coefficients discussed in Chapter IV. A complete description of the EDSTAT-J package is given in Jennings (1971).

In the cross-validation phase of the Policy Capturing process, either the PREDICT routine of the EDSTAT-J package or the NCRIT routine can be used. The NCRIT routine was adapted for use in this research from a program obtained from the AFHRL. It is very similar to the PREDICT routine. It does allow the specification of an arbitrary objective function and operates as a "canned program". A complete listing of NCRIT, including comment cards describing the input, is included.

Program Listings:

In order that a master listing of the programs used in this research could be referenced, source listings for those programs that are not formally described elsewhere are provided. The programs included are:

AID4UT (AIDIV, ERROR, HEADER, LINK3, AIDFIN, RANKSIT)

AIDTRE (TREEC, BLOCK, PLEVEL, PLOTH, PLOTH1, PLOTV, PLOTV1, STOA1, PLOTHI, ITOA1, PLOTHF, RTOA1, AIDPLT)

CRV (The various subroutines referenced are described in Jennings, 1971)

NCRIT (DRIVER, DATRAM, TMT, REGWTS, TABLES, OBJFUN) (This program also uses some EDSTAT-J routines.)

AID4UT

```
      PROGRAM AIDIV(INPUT,OUTPUT,PUNCH,TAPE5=INPUT,TAPE1,TAPE2,TAPE3,TAP
     1E4,TAPE7=PUNCH,TAPE8,TAPE10,TAPE14,TAPE25,TAPE19)
      DIMENSION IFM(52), ICOMP(700), AVER(465), SD(465), LABEL(8), J2(4,
     14), IICR(81,3),AL(4,2)
      COMMON IRUN,ITREE,NG1,XMG1,SDG1,NG2,XMG2,SDG2,ACR(80,3),NOGO
      COMMON /GARB/    GDATA(90,6),SPLIT(700,4),P1,P2,RC,RE,KA,KB,KC,KD,
     1KE,KF,MAXGP,NMIN,JBL,KBL,LBL,MBL,NBL,NP,NL,ISTOR(465),IW,INOUT(84)
     2,IOCT(10),ICR(80,3),KSTOP,IRESID,IFSW,IFSW2,IPHASE,RSUM(4,5)
     3,XM(180),IESS,IRSQ1,IRSQ2
      COMMON /MMM/ LINER
      COMMON/APT/NPRED,NAMI(30),ITRIA(183)
      EQUIVALENCE (ISTOR,AVER,SD)
      DATA KBLANK/6H      /
      DATA ICRIT,IPRED/6H CRIT ,6H PRED /
      DATA AL(1,1),AL(1,2),AL(2,1),AL(2,2),AL(3,1),AL(3,2),AL(4,1),AL(4,
     12)/6H (NORM,6HAL A) ,6H (FORC,6HE B)  ,6H (NORM,6HAL B) ,6H (FORC,
     26HE A)  /
      REAL NAMI
C     **************************************************************
C     * INITIALIZE PARAMETERS                                      *
C     **************************************************************
      IW=6
      JCE=0
C     **************************************************************
C     * READ AID TITLE CARD                                        *
C     **************************************************************
C     * CARD COLUMNS     DESCRIPTION                               *
C     **************************************************************
C     * 1              R  CARD TYPE=1                              *
C     *                   NOTE THAT CARD TYPE=9 DENOTES END OF RUN *
C     * 2-49           O  TITLE                                    *
C     * 50             R  IRUN-THE RUN OPTION                      *
C     *                      0-RUN TOTAL                           *
C     *                      1-RUN A                               *
C     *                      2-RUN A,FORCE B                       *
C     *                      3-RUN A,FORCE B,RUN B,FORCE A         *
C     *                      4-GIVEN A AND B,RUN A,FORCE B         *
C     * 51-56          R* NCPERM-THE NUMBER OF CASES IN THE ORIGINAL *
C     *                         INPUT FILE(OPTIONAL WHEN IRUN=0)   *
C     *                         (DOES NOT APPLY IF IRUN = 4)       *
C     * 57-58          O  NREELS-THE NUMBER OF REELS FOR NCPERM    *
C     * 59-64          R* NSAMA-THE NUMBER OF CASES TO BE SELECTED IN *
C     *                      SAMPLE A(BLANK IF IRUN=0)             *
C     *                      (OPTIONAL IF IRUN = 4)                *
C     * 65-70          R* NSAMB-THE NUMBER OF CASES TO BE SELECTED IN *
C     *                      SAMPLE B(BLANK IF IRUN=0 OR 1)        *
C     *                      (OPTIONAL IF IRUN = 4)                *
C     * 71-78          R* NUM-AN EIGHT DIGIT STARTING RANDOM NUMBER *
C     *                      (BLANK IF IRUN=0 OR 4                 *
C     * 79-80          R  IFMT-THE NUMBER OF FORMAT CARDS TO BE READ IN *
C     *                        NEXT                                *
C     **************************************************************
      READ 1040, IT,LABEL,IRUN,NCPERM,NREELS,NSAMA,NSAMB,NUM,IFMT
C     **************************************************************
C     * EDIT TITLE CARD                                            *
      NCPERM=IABS(NCPERM)
C     **************************************************************
      NREELS=IABS(NREELS)
```

```
      NSAMA=IABS(NSAMA)
      NSAMB=IABS(NSAMB)
      NUM=IABS(NUM)
      IF (IT.EQ.9) GO TO 840
      IF (IT.NE.1) CALL ERROR (1)
      IF (IRUN.EQ.0.OR.IRUN.EQ.4) GO TO 20
      IF (IRUN.GT.4) CALL ERROR (2)
      IF (NCPERM.EQ.0) CALL ERROR (3)
      IF (NSAMA.EQ.0) CALL ERROR (4)
      IF (IRUN.GT.1.AND.NSAMB.EQ.0) CALL ERROR (5)
      IF (NUM.EQ.0) CALL ERROR (6)
      IF (MOD(NUM,17).EQ.0) CALL ERROR (7)
      IF (MOD(NUM,5882353).EQ.0) CALL ERROR (8)
   20 IF (IFMT.EQ.0.OR.IFMT.GT.4) CALL ERROR (9)
C     **************************************************************
C     * PRINT TITLE CARD                                           *
C     **************************************************************
      PRINT 1050
      PRINT 860, IT,LABEL,IRUN,NCPERM,NREELS,NSAMA,NSAMB,NUM,IFMT
C     **************************************************************
C     * READ FORMAT CARD OR CARDS DEPENDING ON IFMT                *
C     **************************************************************
C     * CARD COLUMNS     DESCRIPTION                               *
C     **************************************************************
C     * 1-78           R  IFM-THE FORMAT OF THE INPUT              *
C     *                      (IFM IS DIMENSIONED 52 TO PROCESS A   *
C     *                      MAXIMUM OF FOUR FORMAT CARDS)         *
C     * 79-80             NOT USED                                 *
C     **************************************************************
      IFMT=IFMT*13
      READ 920, (IFM(I),I=1,IFMT)
C     **************************************************************
C     * PRINT FORMAT CARD OR CARDS                                 *
C     **************************************************************
      PRINT 930
      PRINT 900, (IFM(I),I=1,IFMT)
C     **************************************************************
C     * READ DESCRIPTION CARD                                      *
C     **************************************************************
C     * CARD COLUMNS     DESCRIPTION                               *
C     **************************************************************
C     * 1              R  CARD TYPE=3                              *
C     * 2-6            R  P1-THE PROPORTION OF THE TOTAL SUM OF SQUARES *
C     *                      THAT MUST BE CONTAINED IN ANY GROUP IF THAT*
C     *                      GROUP IS TO BECOME A CANDIDATE GROUP FOR *
C     *                      SPLITTING                             *
C     * 7-11           R  P2-THE BEST SPLIT ON THE ITH CANDIDATE GROUP *
C     *                      MUST REDUCE THE UNEXPLAINED SUM OF SQUARES *
C     *                      BY P2 PROPORTION OF THE TOTAL SUM OF  *
C     *                      SQUARES OR THAT GROUP WILL NOT BE SPLIT, *
C     *                      AND IT WILL NOT BECOME A CANDIDATE GROUP *
C     *                      AGAIN EVEN THOUGH IT MAY MEET THE P1  *
C     *                      REQUIREMENT.                          *
C     * 12-16          R  MAXGP-THE MAXIMUM NUMBER OF FINAL GROUPS. *
C     *                         IF FORCED SPLITTING IS CALLED FOR, THIS *
C     *                         NUMBER MAY NOT EXCEED 90           *
C     * 17-21          R  NMIN-THE MINIMUM NUMBER OF CASES IN BOTH *
C     *                        RESULTANT GROUPS REQUIRED FOR A SPLIT *
```

```
C     * 22-26           R* KSTOP-THE NUMBER OF ITERATIONS TO BE PRINTED  *
C     *                        IN DETAIL.AFTER THIS NUMBER IS EXCEEDED.*
C     *                        ONLY SUMMARY INFORMATION ABOUT SPLITS IS*
C     *                        PRINTED.(IF KSTOP IS LEFT BLANK,KSTOP   *
C     *                        WILL BE SET TO 0)                       *
C     *                        IF KSTOP EQUALS -1,ONLY A SPLIT SUMMARY *
C     *                        WILL BE PRINTED                         *
C     * 27-29           R  NV-THE NUMBER OF VARIABLES TO BE READ IN    *
C     *                    INCLUDING THE CRITERION VARIABLE AND THE    *
C     *                    IDENTIFICATION VARIABLE  AND THE WEIGHT     *
C     *                    VARIABLE.NV MUST BE LESS THAN OR EQUAL      *
C     *                    TO 83.                                      *
C     * 30-32           R  NR-THE FIELD NUMBER OF THE IDENTIFICATION   *
C     *                    VARIABLE                                    *
C     * 33              O  KRW-THE REWIND OPTION                       *
C     *                      0-NO REWIND OF INPUT TAPE                 *
C     *                      1-REWIND INPUT TAPE                       *
C     * 34              O  IOPT-THE EXCLUDE OPTION FOR PREDICTOR       *
C     *                         VARIABLES                              *
C     *                      0-DO NOT EXCLUDE CASES WITH OUT-OF-RANGE  *
C     *                        PREDICTOR VALUES.ASSIGN MINIMUM OR      *
C     *                        MAXIMUM ACCORDINGLY.                    *
C     *                      1-EXCLUDE CASES WITH OUT-OF-RANGE PREDICTOR*
C     *                        VALUES                                  *
C     * 35              O  KREJ-THE REJECT SUMMARY PRINT OPTION        *
C     *                      0-NO REJECT SUMMARY TO BE PRINTED         *
C     *                      1-REJECT SUMMARY TO BE PRINTED.           *
C     * 36              O  IRESID-THE RESIDUAL SUMMARY OPTION          *
C     *                        0-NO RESIDUALS                          *
C     *                        1-PRINT RESIDUALS                       *
C     *                        2-PUNCH RESIDUALS                       *
C     *                        (IRUN MUST BE LESS THAN 2)              *
C     *                        3-WRITE RESIDUALS ON TAPE               *
C     *                        (IRUN MUST BE LESS THAN 2)              *
C     * 37              O  ICARD-THE CARD INPUT OPTION                 *
C     *                       0-NO CARD INPUT                          *
C     *                       1-CARD INPUT(IRUN MUST BE LESS THAN 3)   *
C     *                         (INPUT DATA ON TAPE25)                 *
C     * 38              O  ITREE-THE TREE PLOT OPTION                  *
C     *                       0-NO TREE                                *
C                               3-STRUCTURE PLOT ONLY
C                             2-STRUCTURE PLOT AND DETAIL PLOT
C     *                         1-TREE                                 *
C     * 39              O  NOGO-THE NOGO OPTION                        *
C     *                       0-GO                                     *
C     *                       1-NOGO                                   *
C     * 40              O  IESS-OPTION TO PRINT ERROR SUM OF SQUARES IN *
C     *                        FORCED SPLITS(CROSS-VALIDATION).COMPUTED *
C     *                        FROM MEANS OF PREVIOUS SAMPLE.          *
C     *                      0-DO NOT PRINT                            *
C     *                      1-PRINT                                   *
C     * 41              O  IRSQ1-OPTION TO PRINT 1 - ESS(FULL)/ESS(REST.)*
C     *                        IN FORCED SPLITS,COMPUTED FROM MEANS OF *
C     *                        PREVIOUS SAMPLE.                        *
C     *                      0-DO NOT PRINT                            *
C     *                      1-PRINT                                   *
C     * 42              O  IRSQ2-OPTION TO PRINT BETA(TRANSPOSE)*VALIDITY*
C     *                        IN FORCED SPLITS,WITH BETA WEIGHTSFROM  *
C     *                                     PREVIOUS SAMPLE.           *
C     *                                     0-DO NOT PRINT             *
C     *                                     1-PRINT                    *
C     * 43-80                       NOT USED                           *
C     *****************************************************************
      READ 1660, IT,P1,P2,MAXGP,NMIN,KSTOP,NV,NR,KRW,IOPT,KREJ,IRESID,IC
     1ARD,ITREE,NOGO,IESS,IRSQ1,IRSQ2
C     *****************************************************************
C     * EDIT DESCRIPTION CARD                                          *
C     *****************************************************************
      IF(KSTOP .EQ. 0) KSTOP = 0
      NR = IABS(NR)
      KRW=IABS(KRW)
      IOPT=IABS(IOPT)
      KREJ=IABS(KREJ)
      IRESID=IABS(IRESID)
      ICARD=IABS(ICARD)
      ITREE=IABS(ITREE)
      NOGO=IABS(NOGO)
      IF(IESS .NE. 1) IESS = 0
      IF(IRSQ1 .NE. 1) IRSQ1 = 0
      IF(IRSQ2 .NE. 1) IRSQ2 = 0
      IF (IT.NE.3) CALL ERROR (10)
      IF (P1.GT.1.) CALL ERROR (11)
      IF (P2.GT.1.) CALL ERROR (12)
      IF (MAXGP.LT.1) CALL ERROR (13)
      IF (NV.GT.83) CALL ERROR (38)
      IF (KRW.GT.1) CALL ERROR (14)
      IF (IOPT.GT.1) CALL ERROR (15)
      IF (KREJ.GT.1) CALL ERROR (16)
      IF (IRESID.GT.3) CALL ERROR (17)
      IF (ICARD.GT.1) CALL ERROR (41)
      IF (IRUN.GT.2.AND.ICARD.EQ.1) CALL ERROR (42)
      IF (IRUN.GT.1.AND.IRESID.GT.1) CALL ERROR (18)
      IF (MAXGP.GT.90.AND.IRUN.GT.1) MAXGP=90
      IF(NMIN.LT.1)NMIN=1
C     *****************************************************************
C     * PRINT DESCRIPTION CARD                                         *
C     *****************************************************************
      PRINT 910, IT,P1,P2,MAXGP,NMIN,KSTOP,NV,NR,KRW,IOPT,KREJ,IRESID,IC
     1ARD,ITREE,NOGO,IESS,IRSQ1,IRSQ2
C     *****************************************************************
      KAPERM=25
      IF (ICARD.EQ.1) KAPERM=5
      KBPERM=4
      KCPERM=3
      KDPERM=1
      KEPERM=2
      KFPERM=8
      IF (IRESID.EQ.2) KFPERM=7
      KA=KAPERM
      KB=KBPERM
      KC=KCPERM
      KD=KDPERM
      KE=KEPERM
      KF=KFPERM
C     *****************************************************************
      IF (NCPERM.EQ.0) NCPERM=1000000
```

```
      IF (NSAMA.EQ.0) NSAMA=1000000
      IF (NSAMB.EQ.0) NSAMB=1000000
      IF(IRUN.EQ.1)NSAMB=0
      KOPT=IOPT+1
      IOP2=2*IOPT
      IF (ICARD.EQ.1) GO TO 30
      IF (KRW.EQ.1) REWIND KAPERM
   30 NC=NCPERM
      IF (IRUN.EQ.5) NC=NSAMA
      IPHASE=0
      J1=1
      J2X=2
      JUMP=1
      IFSW=2
      IFSW2=2
      IF (IRUN.GT.1) IFSW2=1
      LN=KREJ+1
      IF (IRUN.GT.0.AND.IRUN.LT.4.AND.LN.EQ.1) LN=3
      NSEL=0
      NSELA=0
      NSAB=NSAMA+NSAMB
      RATIO=FLOAT(NSAMA)/FLOAT(NSAB)
      ASSIGN 360 TO IWHAT1
      IF (IRUN.EQ.0.OR.IRUN.EQ.4) ASSIGN 480 TO IWHAT1
      ASSIGN 380 TO IWHAT3
      IF (IRUN.EQ.1) ASSIGN 400 TO IWHAT3
      ASSIGN 630 TO IWHAT4
      IF (IRUN.LE.1.OR.IRUN.EQ.4) ASSIGN 670 TO IWHAT4
      KNUM=NUM
      K1=1
      IF (P2.GT.P1) PRINT 1106
      J=7-I
      DO 40 I=J,6
      K=6-I
   40 IOCT(I)=40**K
      K2=1
      K3=1
      LINER=53
      CALL HEADER(1)
      DO 250 I=1,81
C     ***************************************************************
C     * READ PREDICTOR CARD OR CARDS,THEN CRITERION CARD            *
C     ***************************************************************
C     * PREDICTOR CARD                                              *
C     ***************************************************************
C     * CARD COLUMNS       DESCRIPTION                              *
C     ***************************************************************
C     * 1              R  CARD TYPE=4                               *
C     * 2-19           R  NAME OF PREDICTOR                         *
C     * 20-22          R  JBL-FIELD NUMER OF PREDICTOR VARII 35
C     * 23-24          R  KBL-THE PREDICTOR SPECIFICATION           *
C     *                    00-FREE AND EQUAL INTERVALS              *
C     *                    01-FREE AND SPECIFIED CATEGORIES         *
C     *                    10-MONOTONIC AND EQUAL INTERVALS         *
C     *                    11-MONOTONIC AND SPECIFIED CATEGORIES    *
C     ***************************************************************
C     * EQUAL INTERVALS                                             *
C     * 25-30          R  MINIMUM VALUE                             *
C     * 31-36          R  MAXIMUM VALUE                             *
C     * 37-42          R  INTERVAL LENGTH                           *
C     * 43-45          R  SPECIFIED RECODE(-1 IF NOT USED)          *
C     * 46-51          R  INPUT VALUE TO BE GIVEN RECODE            *
C     * 52-54          R  RECODE(-1 IF NOT USED)                    *
C     * 55-60          R  VALUE                                     *
C     * 61-63          R  RECODE(-1 IF NOT USED)                    *
C     * 64-69          R  VALUE                                     *
C     * 70-80             NOT USED                                  *
C     ***************************************************************
C     * SPECIFIED CATEGORIES                                        *
C     * 25-27          R  RECODE                                    *
C     * 28-33          R  VALUE                                     *
C     * 34-36          R  RECODE                                    *
C     * 37-42          R  VALUE                                     *
C     * 43-45          R  RECODE                                    *
C     * 46-51          R  VALUE                                     *
C     * 52-54          R  RECODE                                    *
C     * 55-60          R  VALUE                                     *
C     * 61-63          R  RECODE                                    *
C     * 64-69          R  VALUE-NOTE THAT THE LAST VALUE TO BE USED MUST*
C     *                         ALWAYS EQUAL 999999.IF MORE THAN 5   *
C     *                         VALUES ARE REQUIRED,ADDITIONAL CARDS *
C     *                         WITH COLUMNS 1-24 DUPLICATED MAY BE  *
C     *                         USED                                 *
C     * 70-80             NOT USED                                  *
C     ***************************************************************
C     * CRITERION CARD                                              *
C     * 1              R  CARD TYPE=5                               *
C     * 2-19           R  NAME OF CRITERION                         *
C     * 20-22          R  JBL-FIELD NUMBER OF CRITERION VARIABLE    *
C     * 23-24          O  KBL-FIELD NUMBER OF WEIGHT                *
C     * 25-30          R  YMAX-THE MAXIMUM ALLOWABLE VALUE OF THE   *
C     *                         CRITERION VARIABLE                  *
C     * 31-36          R  YMIN-THE MINIMUM ALLOWABLE VALUE OF THE   *
C     *                         CRITERION VARIABLE                  *
C     * 37-42          R  MD1-ANY CRITERION VARIABLE EQUAL TO MD1 WILL *
C     *                         BE DELETED                          *
C     * 43-48          R  MD2-ANY CRITERION VARIABLE EQUAL TO MD2 WILL *
C     *                         BE DELETED                          *
C     * 49-80             NOT USED                                  *
C     ***************************************************************
C
C     ***************************************************************
C     * THE IICR(I,J) ARRAY STORES THE FOLLOWING INFORMATION FOR PREDICT
C     *      IICR(I,1) = FIELD NUMBER
C     *      IICR(I,2) = 0 IF EQUAL INTERVAL
C     *                = 1 IF SPECIFIED CATEGORIES
C     *      IICR(I,3) = POINTER TO THE BEGINNING OF PREDICTOR(I) IN THE
C     *                  ICOMP ARRAY
C
C     * THE  ICR(L,K) ARRAY STORES THE FOLLOWING INFORMATION FOR PREDICT
C     *      ICR(L,1) = POINTER TO THE LAST CLASS OF PREDICTOR L IN
C                       THE SPLIT ARRAY
C     *      ICR(L,2) = POINTER TO THE FIRST CLASS OF PREDICTOR L IN
C                       THE SPLIT ARRAY
C     *      ICR(L,3) = 0 IF PREDICTOR(L) IS FREE
C     *               = 1 IF PREDICTOR(L) IS MONOTONIC
C     ***************************************************************
```

```
      READ 880, IT,AA,BB,CC,JBL,KBL,(INOUT(J),J=1,15)
      KBL=IABS(KBL)
      IICR(I,3)=K3
      IF (IT.NE.4) GO TO 260
      IF (JBL.GT.NV) CALL ERROR (19)
      IF (JBL.LT.1) CALL ERROR (20)
      IICR(I,1)=JBL
      ICR(I,3)=KBL/10
      CALL HEADER(3)
      IF(ICR(I,3).EQ.1)GO TO 41
      PRINT 1150
      GO TO 42
41    PRINT 1161
42    IICR(I,2)=KBL-10*ICR(I,3)
      IF (ICR(I,3).LT.0) CALL ERROR (21)
      IF (ICR(I,3).GT.1) CALL ERROR (22)
      IF (IICR(I,2).LT.0) CALL ERROR (23)
      IF (IICR(I,2).GT.1) CALL ERROR (24)
      IF(I.LE.30) NAMI(I)=AA
      ACR(I,1)=AA
      ACR(I,2)=BB
      ACR(I,3)=CC
      ICR(I,2)=K2
      IF (IICR(I,2).EQ.1) GO TO 160
      M=3
C
C     ******************************************************************
C     *
C     * FIND MAXIMUM AND MINIMUM RANGES AND INTERVAL LENGTH FOR EACH PRE
C     *
C     ******************************************************************
C     * IF THE PREDICTOR HAS SPECIFIED CATEGORIES,GO TO STATEMENT 350
C     *
C     * THE INOUT(J) ARRAY STORES THE FOLLOWING INFORMATION FOR THE VARI
C     *     FOR THE PREDICTORS                FOR THE CRITERION
C     *       INOUT(1)= MIN.VALUE              INOUT(1)= YMAX
C     *       INOUT(2)= MAX.VALUE              INOUT(2)= YMIN
C     *       INOUT(3)= INTERVAL               INOUT(3)= FIRST  MISSING DAT
C     *       INOUT(4)= SPECIFIED RECODE       INOUT(4)= SECOND MISSING DAT
C     *       INOUT(5)= INPUT VALUE TO BE
C     *                 GIVEN RECODE VALUE
C     *       INOUT(6)= -1,IF END RECODE
C     *               = SAME AS INOUT(4) IF
C     *                 MORE RECODE TO COME
C     *       INOUT(7)- INOUT(10) REPEATS
C     *         STEPS FOR INOUT(5)-INOUT(6)
C     *
C     ******************************************************************
C
      DO 50 J=1,3
      K=2*J
      INOUT(J)=1000*INOUT(K-1)+INOUT(K)
50    CONTINUE
      IF (INOUT(3).LE.0) CALL ERROR (46)
      IR = INOUT(2)-INOUT(1)
      L = IR/INOUT(3)
      IP = IR - L*INOUT(3)
      L = L-1
      IF (L.LT.0) CALL ERROR (25)
      INOUT(4)=INOUT(7)
      J=0
      LA=INOUT(3)
      LB=INOUT(1)+LA
      GO TO (70,60), KOPT
60    IF (INOUT(1).EQ.0) INOUT(1)=0
      PRINT 990, I,(ACR(I,N),N=1,3),IICR(I,1),INOUT(1)
      LX=LB-1
      PRINT 1000, J,INOUT(1),LX
      LINER=LINER+1
      GO TO 80
70    PRINT 950, I,(ACR(I,N),N=1,3),IICR(I,1),J,LB
80    IF (L.LT.1) GO TO 110
      DO 90 J=1,L
      LC=LB+LA-1
      CALL HEADER (1)
      PRINT 1000, J,LB,LC
90    LB=LC+1
      CALL HEADER (1)
      GO TO (110, 95), KOPT
95    IF(IR .EQ. 0) GO TO 100
      L = L+1
      LC = LB +IR - 1
      PRINT 1000,L,LB,LC
      LINER = LINER + 1
100   PRINT 980, INOUT(2)
      GO TO 120
110   L=L+1
      PRINT 1020, L,LB
120   DO 130 J=7,13,3
      IF (INOUT(J).LT.0) GO TO 140
      K=2*J/3
      INOUT(K-1)=1000*INOUT(J+1)+INOUT(J+2)
      INOUT(K)=INOUT(J)
      INOUT(K+2)=-1
      CALL HEADER (1)
      PRINT 1030, INOUT(K),INOUT(K+1)
      M=M+2
130   CONTINUE
C     ******************************************************************
C     * THE ICOMP ARRAY STORES THE FOLLOWING
C     * LET K = IICR(I,3)
C     *FOR EQUAL INTERVALS
C     * ICOMP(K)   = MINIMUM VALUE FOR PREDICTOR I
C     * ICOMP(K+1) = MAXIMUM VALUE
C     * ICOMP(K+2) = INTERVAL LENGTH
C     * ICOMP(K+3) = RECODE VALUE(OUTPUT) FOR NEXT INPUT VALUE
C     * ICOMP(K+4) = INPUT VALUE
C     * ICOMP(K+5) = RECODE VALUE
C     * ICOMP(K+6) = INPUT  VALUE
C     * ICOMP(K+7) = RECODE VALUE
C     * ICOMP(K+8) = INPUT  VALUE
C     * ICOMP(K+9) = -1(NOTE THAT A MAXIMUM OF THREE RECODE VALUES
C     *              ARE PROVIDED.IF LESS THAN THREE,-1 MOVES
C     *              ACCORDINGLY.EXAMPLE-IF NO RECODE VALUES THEN
C     *              ICOMP(K+3)=-1)
C     *FOR SPECIFIED INTERVALS
```

```
C     * FOR IOPT = 1(DISCARD CASES WITH OUT-OF-RANGE PREDICTORS)
C     *   ICOMP(K  ) = 0
C     *   ICOMP(K+1) = LOWER BOUND(DISCARD CASES WITH PREDICTOR I LESS
C     *                THAN OR EQUAL TO LOWER BOUND)
C     *   ICOMP(K+2) = RECODE VALUE(OUTPUT) FOR ALL VALUES LESS THAN OR
C     *                EQUAL TO NEXT INPUT VALUE(ICOMP(K+3))
C     *   ICOMP(K+3) = INPUT VALUE
C     *   ICOMP(K+4) = RECODE VALUE FOR ALL VALUES LESS THAN OR EQUAL
C     *                TO NEXT INPUT VALUE BUT GREATER THAN PREVIOUS
C     *                INPUT VALUE
C     *   ICOMP(K+5) = INPUT VALUE
C     *        .
C     *        .
C     *   ICOMP(K+J) = RECODE VALUE
C     *   ICOMP(K+J+1) = UPPER BOUND(DISCARD CASES WITH PREDICTOR I
C     *                  GREATER THAN UPPER BOUND)
C     * FOR IOPT = 0(DO NOT DISCARD CASES WITH OUT-OF-RANGE PREDICTORS)
C     *   ICOMP(K  ) = RECODE VALUE(OUTPUT) FOR ALL VALUES LESS THAN OR
C     *                EQUAL TO NEXT INPUT VALUE(ICOMP(K+1))
C     *   ICOMP(K+1) = INPUT VALUE
C     *   ICOMP(K+2) = RECODE VALUE FOR ALL VALUES LESS THAN OR EQUAL
C     *                TO NEXT INPUT VALUE BUT GREATER THAN PREVIOUS
C     *                INPUT VALUE
C     *   ICOMP(K+3) = INPUT VALUE
C     *        .
C     *        .
C     *   ICOMP(K+J) = RECODE VALUE FOR ALL VALUES GREATER THAN PREVIOUS
C     *                INPUT VALUE
C     *   ICOMP(K+J+1)=999999
C     ****************************************************************
  140 M=M+1
      DO 150 J=1,M
      K=K3+J-1
      IF (K.GT.700) CALL ERROR (26)
      ICOMP(K)=INOUT(J)
  150 CONTINUE
      GO TO 240
C     ****************************************************************
C     *
C     * THE PREDICTOR WILL HAVE SPECIFIED CATEGORIES
C     *
C     ****************************************************************
  160 L=0
      M=0
      K=K3
      LB=1000*INOUT(2)+INOUT(3)
      IF(LB.EQ.0)LB=0
      GO TO (180,170), KOPT
  170 PRINT 970, I,(ACR(I,N),N=1,3),IICR(I,1),LB
      GO TO 190
  180 J=INOUT(1)
      PRINT 960, I,(ACR(I,N),N=1,3),IICR(I,1),J,LB
  190 DO 210 J=1,13,3
      IF (INOUT(J).GT.L) L=INOUT(J)
      IF (K.GT.699) CALL ERROR (27)
      IANY=1000*INOUT(J+1)+INOUT(J+2)
      IF (IANY.EQ.999999) GO TO (230,220), KOPT
      ICOMP(K)=INOUT(J)
      ICOMP(K+1)=IANY
      M=M+2
      LC=ICOMP(K+1)
      IF (K.EQ.K3) GO TO 200
      CALL HEADER (1)
      PRINT 1000, ICOMP(K),LB,LC
  200 LB=LC+1
      K=K+2
  210 CONTINUE
C     ****************************************************************
C     *
C     * READ CONTINUATION CARDS FOR PREDICTORS WITH SPECIFIED CATEGORIES
C     * UNTILL THE LAST INPUT RECODE VALUE( MANDATORY 999999) IS READ
C     *
C     ****************************************************************
      READ 880, IT,AA,BB,CC,JBL,KBL,(INOUT(J),J=1,15)
      IF (IT.NE.4) CALL ERROR (28)
      IF (JBL.NE.IICR(I,1)) CALL ERROR (29)
      GO TO 190
  220 PRINT 980, LB
      CALL HEADER (1)
      GO TO 240
  230 ICOMP(K)=INOUT(J)
      ICOMP(K+1)=999999
      M=M-2
      PRINT 1020, ICOMP(K),LB
      CALL HEADER (1)
  240 CONTINUE
      K2=K2+L
      IF (L.GT.39) CALL ERROR (30)
      IF (K2.GT.700) CALL ERROR (31)
      ICR(I,1)=K2
      K3=K3+M
      K2=K2+1
  250 CONTINUE
      CALL ERROR (32)
  260 IF (IT.NE.5) CALL ERROR (32)
      IF (JBL.GT.NV) CALL ERROR (33)
      IF (JBL.LT.1) CALL ERROR (34)
      IF (KBL.GT.NV) CALL ERROR (35)
      IF (KBL.LT.0) CALL ERROR (36)
      DO 270 J=1,4
      K=2*J
      INOUT(J)=1000*INOUT(K-1)+INOUT(K)
  270 CONTINUE
      YMAX=INOUT(1)
      YMIN=INOUT(2)
      IF(INOUT(1) .EQ. 0 ) INOUT(1) = 0
      IF(INOUT(2) .EQ. 0 ) INOUT(2) = 0
      IF (INOUT(3).EQ.0) INOUT(3)=0
      IF (INOUT(4).EQ.0) INOUT(4)=0
      Y1=INOUT(3)
      Y2=INOUT(4)
      IF (INOUT(1).LE.INOUT(2)) CALL ERROR (37)
      KL=JBL
      KM=IABS(KBL)
      PRINT 940, JBL,AA,BB,CC
      INOUT(84)=1
```

```
      PRINT 1010, (INOUT(J),J=1,4),KM
      IF(NOGO.EQ.1)  CALL EXIT
      IF (KM.EQ.0) KM=84
      NP=I-1
      NW=NP/IW
      IF (IW*NW.LT.NP) NW=NW+1
      NL=NW
      IF (IRESID.GT.0) NL=NL+1
      IBL=155/(NL+3)
      JBL=2*IBL
      NBL=JBL+1
      KBL=IBL*(NL+3)
      LBL=KBL+155
      MBL=KBL+310
280   DO 300 I=1,90
      DO 290 J=1,6
290   GDATA(I,J)=0.0
300   CONTINUE
      DO 320 I=1,700
      DO 310 J=1,6
310   SPLIT(I,J)=0.0
320   CONTINUE
      REWIND KB
      REWIND KC
      REWIND KO
      REWIND KE
      IF((IRUN.EQ.2.OR.IRUN.EQ.3.OR.(IRUN.EQ.4.AND.KRW.EQ.1))REWIND KF
      IRLCT=0
      NIN=0
      KCNT=0
      NOUT=0
      LMA=NBL
      LMB=NBL+155
      LM=NBL
      LX=1
      LXA=1
      LXB=156
330   NIN=NIN+1
      IF (NIN.GT.NC) GO TO 710
C     ****************************************************************
C     *
C     *            READ  ONE  RECORD  FROM  INPUT  DEVICE
C     ****************************************************************
C     *
C     *
C     *     KM = FIELD NO. OF WEIGHT
C     *     W = VALUE OF  WEIGHT
C     *
C     *     LN = 1 IF NO REJECTS PRINTED,NO RANDOM SAMPLING
C     *          2 IF REJECTS PRINTED
C     *          3 IF NO REJECTS PRINTED,RANDOM SAMPLING
C     *     KL = FIELD NUMBER OF THE VARIABLE OF CRITERION
C     *     M  = VALUE OF THE VARIABLE(CRITERION)
C     *     LK = POINTER FOR THE ICOMP(J) ARRAY
C     *
C     IWHAT1=1ST ARG,RANDOM SAMPLING
C     IWHAT1=2ND ARG,NO RANDOM SAMPLING
C     ****************************************************************
      GO TO IWHAT1, (360,400)
340   READ (KA,IFM)(IXS,J=1,NV)
      GO TO 330
350   NSEL=NSEL-1
      IF (JUMP.EQ.J1) NSELA=NSELA-1
      GO TO 330
360   KNUM=MOD(23*KNUM,100000001)
      RNUM=FLOAT(KNUM)*9.9999999E-9
      PROI=FLOAT(NSAB-NSEL)/FLOAT(NC-NIN+1)
      IF (RNUM.GE.PROI) GO TO 340
      NSEL=NSEL+1
      GO TO IWHAT3, (380,600)
380   PROB=(FLOAT(NSAMA-NSELA)-FLOAT(NOUT)*RATIO)/FLOAT(NC-NIN+1)
      IF (RNUM.GE.PROB) GO TO 390
      NSELA=NSELA+1
      JUMP=J1
      LM=LMA
      LX=LXA
      GO TO 400
390   JUMP=J2X
      LM=LMB
      LX=LXB
400   READ (KA,IFM) (INOUT(J),J=1,NV)
      IF (INOUT(KM).LE.0) INOUT(KM)=1
      W=INOUT(KM)
      ISTOR(LM)=W
      A=INOUT(KL)
      IF (A.GT.YMAX) GO TO (480,460,480), LN
      IF (A.LT.YMIN) GO TO (480,460,480), LN
      IF (A.EQ.-1) GO TO (480,460,480), LN
      IF (A.EQ.-2) GO TO (480,460,489), LN
      IF (IOPT) 490,490,410
C     ****************************************************************
C     CHECK FOR OUT OF RANGE PREDICTORS
410   IVAL=1
420   LVAL=IICR(IVAL,1)
      MVAL=INOUT(LVAL)
      LKVAL=IICR(IVAL,3)
      IF (IICR(IVAL,2).EQ.0) GO TO 430
C     SPECIFIED INTERVALS
      IF (ICOMP(LKVAL+1).GE.MVAL) GO TO (480,450,480), LN
      LKVAL=IICR(IVAL+1,3)-1
      IF (ICOMP(LKVAL).LT.MVAL) GO TO (480,450,480), LN
      GO TO 440
C     EQUAL INTERVALS
430   IF (MVAL.GE.ICOMP(LKVAL+1)) GO TO (480,450,480), LN
      IF (MVAL.LT.ICOMP(LKVAL)) GO TO (480,450,480), LN
440   IVAL=IVAL+1
      IF (IVAL.LE NP) GO TO 420
      GO TO 490
450   KZ=IPRED
      IK=IVAL
      GO TO 470
C     ****************************************************************
C     *
C     *     PRODUCE REJECTS   FOR MISSING DATA CASES
C     *
C     ****************************************************************
```

```
468   IK=INOUT(KL)
      K2=ICRIT
470   JOE=JOE+1
      J2(JOE,1)=K2
      J2(JOE,2)=INOUT(NR)
      IF(NR .EQ. 0) J2(JOE,2) = 0
      J2(JOE,3)=INOUT(KN)
      J2(JOE,4)=IK
      IF (JOE.LT.4) GO TO 480
      IF (MOD(NOUT,200).EQ.3) PRINT 1080
      PRINT 1090, ((J2(II,KK),KK=1,4),II=1,4)
      JOE=0
480   CONTINUE
      NOUT=NOUT+1
      IF (IRUN.EQ.0.OR.IRUN.EQ.4) GO TO 330
      GO TO 350
490   Y=A*W
      LM=LM+1
      IF (IRESID.EQ.0) GO TO 500
      ISTOR(LM)=INOUT(NR)
      LM=LM+1
500   YY=A*Y
      I=1
C     ****************************************************************
C     CODE RESPONSES BASE 40
      DO 600 J=1,NW
      IWORK=0
      DO 590 K=1,IW
      IF (I.GT.NP) GO TO 520
      L=IICR(I,1)
      M=INOUT(L)
      LK=IICR(I,3)
      IF (IICR(I,2).EQ.1) GO TO 530
C     ****************************************************************
C     *
C     *               PREDICTORS WITH EQUAL INTERVAL
C     *
C     ****************************************************************
      IF (M.GT.ICOMP(LK+1)) M=ICOMP(LK+1)
      IF (M.LT.ICOMP(LK)) M=ICOMP(LK)
      LL=(M-ICOMP(LK))/ICOMP(LK+2)
510   IF (ICOMP(LK+3).LT.0) GO TO 560
      IF (        M.EQ.ICOMP(LK+4)) LL=ICOMP(LK+3)
      LK=LK+2
      GO TO 510
520   LL=0
      GO TO 580
C     ****************************************************************
C     *
C     *               PREDICTORS WITH SPECIFIED CATEGORIES
C     *
C     *               LL= VALUE TO BE GIVEN TO THE CATEGORY
C     *
C     ****************************************************************
530   K3=IICR(I,3)+IOP2
      K4=IICR(I+1,3)-3
      DO 540 L=K3,K4,2
      IF (M.LE.ICOMP(L+1)) GO TO 550
540   CONTINUE
      L=K4+1
550   LL=ICOMP(L)
560   GO TO (570,580), JUMP
570   M=ICR(I,2)+LL
      SPLIT(M,1)=SPLIT(M,1)+1.0
      SPLIT(M,2)=SPLIT(M,2)+W
      SPLIT(M,3)=SPLIT(M,3)+Y
      SPLIT(M,4)=SPLIT(M,4)+YY
580   IWORK=40*IWORK+LL
      I=I+1
590   CONTINUE
      ISTOR(LM)=IWORK
600   LM=LM+1
C     ****************************************************************
      GO TO (610,620), JUMP
610   KCNT=KCNT+1
      GDATA(1,2)=GDATA(1,2)+1.0
      GDATA(1,3)=GDATA(1,3)+W
      GDATA(1,4)=GDATA(1,4)+Y
      GDATA(1,5)=GDATA(1,5)+YY
620   AVEH(LX)=Y
      SC(LX+1)=YY
      GO TO IWHAT4, (630,670)
630   IF (JUMP.EQ.3)) GO TO 650
      IF (LM.GE.LBL) GO TO 640
      LXB=LXB+2
      LMB=LM
      GO TO 330
640   WRITE (KF) K1,(ISTOR(J),J=156,LBL)
      LMB=NBL+155
      LXB=156
      GO TO 330
650   IF (LM.GE.MAL) GO TO 660
      LAA=LAA+2
      LMA=LM
      GO TO 330
660   LMA=NAL
      LAA=1
      GO TO 690
670   IF (LM.GE.KBL) GO TO 680
      LX=LX+2
      GO TO 330
680   LM=NAL
      LX=1
690   WRITE (KC) K1,(ISTOR(J),J=1,KBL)
      GO TO 330
700   CONTINUE
      INLCT=INLCT+1
710   IF(NOUT.LT.4.AND.JOE.NE.0)PRINT 1080
      IF (JOE.NE.0) PRINT 1090, ((J2(II,KK),KK=1,4),II=1,JOE)
      IF(IPHASE.EQ.2)GO TO 711
      NEN=KCNT
      IF(IRUN.EQ.2.OR.IRUN.EQ.3)NEN=NSEL
      BP=(FLOAT(NOUT)/FLOAT(NOUT+NEN))*100.
      IF(HP.LT..0000I)GO TO 711
      PRINT 1162,HP
711   IF (IRUN.LT.2.OR.IRUN.EQ.4) GO TO 730
```

```
      IF (LKB.EQ.156) GO TO 720
      ISTOR(LMB)=999999
      WRITE (KF) KI,(ISTOR(J),J=156,LBL)
726   LL=999999
      WRITE (KF) LL,(ISTOR(J),J=156,LBL)
      LM=LMA
      LX=LXA
739   IF (LX.EQ.1) GO TO 740
      ISTOR(LM)=999999
      WRITE (KC) KI,(ISTOR(J),J=1,KBL)
740   LL=999999
      WRITE (KC) LL,(ISTOR(J),J=1,KBL)
      GDATA(1,1)=KI
      REWIND KC
      IF (IRUN.GT.1.AND.IRUN.LT.4) REWIND KF
      A=GDATA(1,4)/GDATA(1,3)
      GDATA(1,6)=GDATA(1,5)-A*GDATA(1,4)
      S=GDATA(1,6)/GDATA(1,3)
      NT=GDATA(1,2)
      S=SQRT(S)
      RC=P2*GDATA(1,6)
      RE=P1*GDATA(1,6)
      IF(IRUN.EQ.4.AND.IPHASE.EQ.1)GO TO 747
      PRINT 890, RE,RC,KCNT
747   PRINT 870
      PRINT 850, KI,NT,(GDATA(1,J),J=3,6),A,S
      IF(ITREE.EQ.0.OR.IPHASE.EQ.1)GO TO 749
      IF(IPHASE.EQ.2)GO TO 748
      NG1=NT
      XMG1=A
      SDG1=S
      GO TO 749
  748 NG2=NT
      XMG2=A
      SDG2=S
749   KA=KB
      KB=KE
      NC=KCNT
      I=IRUN+1
      GO TO (750,750,780,790,830), I
750   CALL LINK3
760   PRINT 1110
      NPRED=NP
      IF(IRUN.LE.1) CALL AIDFIN
      IF(ITREE.GE.1) WRITE (14) IRUN, ITREE,NG1,XMG1,SDG1,NG2,XMG2,SDG2
     1,ACR,NOGO
      CALL EXIT
  770 PRINT 1150,(I,AL(I,1),AL(I,2),(RSUM(I,J),J=1,5),I=1,IPHASE)
      GO TO 760
780   CALL LINK3
      IPHASE=IPHASE+1
      IF (IPHASE.EQ.2) GO TO 770
      IDUM=IPHASE+1
      PRINT 1120, IPHASE
      PRINT 1140, IDUM
      PRINT 1130
      IFS=1
      IFS2=2
      GO TO 780
799   IF (IPHASE.NE.2) GO TO 800
      L=KC
      KC=KF
      KF=L
800   CALL LINK3
      IPHASE=IPHASE+1
      IDUM=IPHASE+1
      PRINT 1120, IPHASE
      GO TO (810,820,810,770), IPHASE
810   IFS=1
      IFS2=2
      PRINT 1140, IDUM
      PRINT 1130
      GO TO 800
820   IFS=2
      IFS2=1
      PRINT 1140, IDUM
      LA=3
      JOE=0
      KI=1
      NC=NCPERM
C     ******************************************************************
      KA=KAPERM
      KB=KBPERM
      KC=KCPERM
      KD=KDPERM
      KE=KEPERM
      KF=KFPERM
      REWIND KA
C     ******************************************************************
      J1=2
      J2=1
      KNUM=NUM
      NSEL=0
      NSELA=0
      GO TO 280
830   CALL LINK3
      IPHASE=IPHASE+1
      IF (IPHASE.EQ.2) GO TO 770
      IDUM=IPHASE+1
      PRINT 1120, IPHASE
      PRINT 1140, IDUM
      PRINT 1130
      IFS2=2
      KA=KAPERM
      KB=KBPERM
      KC=KCPERM
      KD=KDPERM
      KE=KEPERM
      NC=NSAMB
      JOE=0
      GO TO 280
840   PRINT 1070
      STOP
C
850   FORMAT (2I7,6G17.8)
860   FORMAT (2X,I1,3X,8A6,4X,I1,3X,I6,4X,I2,4X,I6,2X,I6,2X,I8,3X,I2)
```

```
870   FORMAT (/8H   GROUP,4X,1HN,5X,12HTOTAL WEIGHT,8X,8HSUM OF Y,6X,12H
     1SUM Y-SQUARE,10X,5HT S S,12X,4HMEAN,16X,9HSTD. DEV.)
880   FORMAT (I1,3A6,I3,I2,18I3)
890   FORMAT (34H1DO NOT SPLIT UNLESS TSS MORE THAN,G20.8,18H AND BSS MO
     1RE THAN,G20.8,3H.  /8X,I7,1X,32HVALID CASES. GROUP 1 VALUES ARE-)
900   FORMAT (1X,13A6)
910   FORMAT (/*0DESCRIPTION CARD*//* CARD*/* TYPE      P1          P2    M
     1AXGF   NMIN  KSTOP   NV   NR  KRW  IOPT  KREJ  IRESID  ICARD  ITRE
     2E  NOGO  IESS  IRSQ1  IRSQ2*/2X,I
     31,3X,F8.5,2X,F8.5,2X,I5,2X,I5,2X,I5,2X,I3,2X,I3,3X,I1,5X,I1,5X,I1,
     45X,I1,6X,I1,6X,I1,5X,I1,5X,I1,2(6X,I1))
920   FORMAT (13A6)
930   FORMAT (/* FORMAT CARD OR CARDS*//)
940   FORMAT (//34H DEPENDENT VARIABLE-Y IN FIELD NO.,I4,5H IS  ,3A6/)
950   FORMAT ( 1X,I2,2X,3A6,I4,I9,4X,9HLESS THAN,I7)
960   FORMAT ( 1X,I2,2X,3A6,I4,I9,2X,13HLT. OR EQ. TO,I5)
970   FORMAT ( 1X,I2,2X,3A6,I4,*  EXCLUDE  LT. OR EQ. TO*,I5)
980   FORMAT (30X,*EXCLUDE*,I9,* OR OVER*)
990   FORMAT ( 1X,I2,2X,3A6,I4,*  EXCLUDE    LESS THAN*,I7)
1000  FORMAT (30X,I6,I10,3H TO,I7)
1010  FORMAT (4X,8HYMAX. IS,I7,11H...YMIN. IS,I7,10H...EXCLUDE,2I7,1H,/4
     1X,19HWEIGHT IN FIELD NO.,I4//)
1020  FORMAT (30X,I6,I10,8H OR OVER)
1030  FORMAT (30X,I6,I12)
1040  FORMAT (I1,8A6,I1,I6,I2,2I6,I8,I2)
 1050 FORMAT(*1UNIVERSITY OF TEXAS VERSION OF AID4 AS ADAPTED FROM AFHRL
     1,LACKLAND AFB*/* MODIFIED FOR CDC6600 BY CAPT LL GOOCH, OCT 72*//*
     2CONTROL CARD PARAMETERS*//* TITLE CARD*//* CARD*/* TYPE
     3             TITLE                              IRUN  NCPERM  NREELS  NSAM
     4A   NSAMB      NUM     IFMT*)
1060  FORMAT (I1,2F5.5,3I5,2I3,10I1)
1070  FORMAT (*1END OF JOB*)
1080  FORMAT (1H1,58X,*REJECT SUMMARY*//4(* REASON          ID     WT  VALU
     1E*,2H**)/4(31X,2H**)/4(31X,2H**))
1090  FORMAT (1X,4(A6,I10,I6,I7,3H ** ))
1100  FORMAT (//* NOTE THAT P2 IS GREATER THAN P1 *//)
1110  FORMAT (//* END OF AID4UT - DATA FOR TREE ON TAPES 10 + 14*)
1120  FORMAT (//* END OF PHASE *,I2)
1130  FORMAT (//* FORCED SPLITS*//)
1140  FORMAT (*1START OF PHASE *,I2//)
 1150 FORMAT(*1*,37X,*CROSS-VALIDATION TABLE*///* PHASE*,21X,*R-SQUARED*
     1,16X,*R*,16X,*F-ANOVA*,7X,*DF1*,7X,*DF2* /(1H0,I3,2A6,F20.8,F20.5,
     2F20.4,2F10.0))
1160  FORMAT(1H0,60X,*FREE-FLOATING*)
1161  FORMAT(1H0,60X,*MONOTONIC*)
1162  FORMAT(//* PERCENTAGE OF EXAMINED CASES REJECTED = *,F8.2,* PERCEN
     1T*)
      END


      SUBROUTINE ERROR (IX)
      PRINT 10, IX
      STOP
C
10    FORMAT (//* CONTROL CARD ERROR *,I3,*,TERMINATE*)
      END
```

```
$IBFTC HEADER
      SUBROUTINE HEADER (I)
      COMMON /MMM/ LINER
      LINER=LINER+I
      IF (LINER.LT.52) GO TO 10
      PRINT 20
      LINER=1
10    RETURN
C
20    FORMAT (1H1,4X,*VARIABLE*,8X,*FIELD NO.*,2X,*RECODE*,6X,*CORRESPON
     1DS TO*,5X,*TYPE*//)
      END
```

```
      SUBROUTINE LINK3
C     ****************************************************************
C     SUMMARY OF SWITCHES AND ARRAYS
C     IFSW  -1 IF FORCED SPLITTING
C           -2 IF REGULAR SPLITTING
C     IFSW1 -1 IF FORCED SPLITTING
C           -2 IF REGULAR SPLITTING
C     IFSW2 -1 IF REGULAR SPLITTING,BUT PREPARING FOR FORCED
C           -2 IF *REGULAR* REGULAR SPLITTING OR FORCED SPLITTING
C     IPHASE-0 TIME 1(REGULAR)
C           -1 TIME 2(FORCED)
C           -2 TIME 3(REGULAR)
C           -3 TIME 4(FORCED)
C     GDATA(90,    1           2          3          4         5          6)
C       GROUP NUMBER           N        SUM W      SUM Y  SUM Y**2       TSS
C     SPLIT(700,   1           2          3          4)
C                  N     SUM W        SUM Y  SUM Y**2
C     ISS1  -1 IF LOWER SPLIT IS OKAY(CASES PRESENT)
C     ISS1  -2 IF LOWER SPLIT IS ZERO
C     ISS2  -1 IF ABOVE SPLIT IS OKAY(CASES PRESENT)
C     ISS2  -2 IF ABOVE SPLIT IS ZERO
C     ****************************************************************
      DIMENSION AVER(465), SD(50), J22(4,4), IFORCE(90,8), MAMA(180,2),
     1INVIN(40), ICODE(40)
      COMMON IRUN,ITREE,NG1,XMG1,SDG1,NG2,XMG2,SDG2,ACR(80,3),NOGO
      COMMON /GARB/    GDATA(90,6),SPLIT(700,4),P1,P2,RC,RE,KA,KB,KC,KD,
     1KE,KF,MAXGP,NMIN,JBL,KBL,LBL,MBL,NBL,NP,NL,ISTOR(465),IW,INOUT(86)
     2,IOCT(10),ICR(80,3),KSTOP,IRESID,IFSW,IFSW2,IPHASE,RSUM(4,5)
     3,XM(180),IESS,IRSQ1,IRSQ2
      COMMON/APT/NPRED,NAMI(30),ITRIA(183)
      EQUIVALENCE (ISTOR,AVER)
C     ****************************************************************
      DIMENSION BETA(180)
      RWVDOT = 0.
      BETA(1) = 0.
      LUMOPT = IESS + IRSQ1 + IRSQ2
C     ****************************************************************
      NPRED=NP
      ISS1=1
      ISS2=1
      LC = 100
      IFC=0
      C=0
      KENDSW=1
      IA1=1
      IF (IFSW2.NE.1) GO TO 20
      DO 10 I=1,186
10    MAMA(I,2)=2
20    IFSW1=2
      IF (IPHASE.EQ.1) IFSW1=1
      ITIME=2
      IDF=1
      TK1=0
      TBSS=0
      IV=7-IW
      N1=2
      IF (IRESID.EQ.0) N1=1
      L2=1
      L3=1
      K1=1
      GO TO (30,40), IFSW
30    IB=1
      IA=2
      KC=KF
      K4=1
      GO TO 910
40    TTSS=GDATA(1,6)
      WRITE (19,9880) TTSS
      IDIOT=0
      GOOGOLS=10.**40
      IF(IRUN.NE.4.OR.IPHASE.NE.1)GO TO 41
      ACC1=XM(1)*GDATA(1,4)
      ACC2=XM(1)*GDATA(1,3)
      ACC3=(XM(1)**2)*GDATA(1,3)
      PRINT 3005,ACC1,ACC2,ACC3
 3005 FORMAT(//* TRACE 3 *,3E16.5//)
41    TNN=GDATA(1,2)
      TNV=GDATA(1,3)
      TSY1S=GDATA(1,4)*GDATA(1,4)/GDATA(1,3)
      GD14=GDATA(1,4)
      GD15=GDATA(1,5)
      IF(IFSW2.EQ.1)XM(1)=GD14/TNV
50    REWIND KA
      IF (L3.GE.MAXGP.OR.L2.GE.90) GO TO 330
      IF(KSTOP.LT.0)GO TO 70
      IF (IFSW1.EQ.1) GO TO 60
      IF (L3.GT.KSTOP) GO TO 70
60    PRINT 1440, L3,K1
70    CONTINUE
      GO TO (420,30), IFSW)
80    DO 90 J=1,NO
90    INOUT(J)=0
      WSS=0.0
      NT=GDATA(1,2)
      W=GDATA(1,3)
      Y=GDATA(1,4)
      AT=Y/W
C
C     ****************************************************************
C     *
C     * COMPUTE TEST STATISTICS ONE VARIABLE AT A TIME
C     *
C     *  K1 = 1,....,C1          FOR THE FIRST  VARIABLE               *
C     *     = (C1+1),....,C2     FOR THE SECOND VARIABLE               *
C     *     = (C2+1),....,C3     FOR THE THIRD  VARIABLE               *
C     *                   ETC.                                         *
C     *      WHERE  C1=M2 = NUMBER OF CLASSES IN THE FIRST  VARIABLE   *
C     *             C2=M2 = NUMBER OF CLASSES IN THE SECOND VARIABLE   *
C     *             C3=M2 = NUMBER OF CLASSES IN THE THIRD  VARIABLE   *
C     *
C     ****************************************************************
C
      DO 260 I=1,NP
      IF (L3.GT.KSTOP) GO TO 100
      PRINT 1450, I,(ACR(I,J),J=1,3)
100   CONTINUE
```

```
      DO 110 J=1,40
110   AVER(J)=40**15
      M1=ICR(I,2)
      M2=ICR(I,1)
C
C     *****************************************************************
C     *
C     * COMPUTE CLASS STATISTICS ONE CLASS AT A TIME
C     *
C     *****************************************************************
C
      K=0
      M3=-1
      DO 170 J=M1,M2
      N=SPLIT(J,1)
      IF (N.EQ.0) GO TO 170
      A=SPLIT(J,3)/SPLIT(J,2)
      IF (ICR(I,3).EQ.1) GO TO 150
C
C     *****************************************************************
C     *
C     * IF PREDICTOR IS MONOTONIC(ICR(I,3)=1),GO TO 1880
C     * IF PREDIC.OR IS FREE FLOATING,SORT CLASS MEANS INTO ASCENDING OR
C     *
C     *  A = CLASS MEAN                                                 *
C     *  AVER(K)= 40**5   FOR ALL K INITIALLY                           *
C     *         = LARGEST CLASS MEAN FOR CURRENT  K                     *
C     *         = 40**5   IF K  IS INCREMENTED IN DO LOOP 1060          *
C     *
C     *****************************************************************
C
      DO 120 K=1,39
      IF (A.LT.AVER(K)) GO TO 130
120   CONTINUE
130   IF (AVER(K).EQ.AVER(40)) GO TO 160
      L=40-K
      DO 140 M=1,L
      MM=41-M
      AVER(MM)=AVER(MM-1)
      SD(MM)=SD(MM-1)
      ICODE(MM)=ICODE(MM-1)
140   CONTINUE
      GO TO 160
150   K=K+1
160   ICODE(K)=IABS(J-M1)
      M3=M3+1
      AVER(K)=A
      SD(K)=SPLIT(J,4)/SPLIT(J,2)-A*A
      SD(K)=SQRT(SD(K))
170   CONTINUE
      IF (L3.GT.KSTOP) GO TO 180
C     *****************************************************************
C     *
C     * WRITE HEADING FOR CLASS STATISTICS
C     *
C     *****************************************************************
      PRINT 1430
180   CONTINUE
```

```
      W1=0.0
      Y1=0.0
      B=0.0
      M4=0
      NN=0
C
C     *****************************************************************
C     *
C     * SEARCH FOR THE LARGEST BSS AMONG THE CLASSES OF THIS VARIABLE.
C     *
C     *****************************************************************
C     * M3 = NUMBER OF CLASSES OF THIS VARIABLE
C     * M1 = SAME AS DEFINED IN DO LOOP 1100  ABOVE                     *
C     * J  = 1,2,...,NV
C     *
C     * Y1 = SUM OF WEIGHTED CRITERION SCORES FOR THIS CLASS(K)
C     * N  = NUMBER OF OBSERVATIONS IN THIS CLASS(K)
C     * W1 = SUM OF WEIGHTS FOR THIS CLASS(K)
C     *
C     * Y  = TOTAL SUM OF WEIGHTED CRITERION SCORES IN ALL CLASSES(TOTAL
C     * NN = NUMBER OF OBSERVATIONS SUMMED OVER CLASSES OF THIS VARIABLE
C     * W  = TOTAL SUM OF WEIGHTS FOR ALL CLASSES OF THIS VARIABLE      *
C     *
C     * Y2 = SUM OF WEIGHTED CRITERION SCORES FOR THE (TOTAL - K) CLASSE
C     * W2 = SUM OF WEIGHTS OF (ALL-K) CLASSES OF THIS VARIABLE         *
C     *
C     * S  = BETWEEN SUM OF SQUARES(BSS) =(Y1**2)/W1 +(Y2**2)/(W-W1)-(Y*
C     *
C     * B  = CURRENT LARGEST BSS.
C     * M4 = POINTER SET TO THE CLASS(K) WITH THE LARGEST CURRENT BSS
C     *
C     *      AFTER EACH CYCLE OF THIS LOOP,THE FOLLOWING ARE PRINTED
C     *               PRINT                       INTERNAL SYMBOL
C     *
C     *               CODE                        ICODE(K)
C     *               N                           N
C     *               TOTAL WEIGHT                SPLIT(J,2)
C     *               SUM OF Y                    SPLIT(J,3)
C     *               SUM Y-SQUARE                SPLIT(J,4)
C     *               MSS                         S
C     *               MEAN                        AVER(K)
C     *               STD.DEV.                    SD(K)
C     *
C     * IF THE NUMBER OF OBSERVATIONS IN THE RESULTANT GROUPS AFTER     *
C     * THE ATTEMPTED SPLIT IS LESS THAN THE MINIMUM SPECIFIED ON CARD *
C     * NO.3,COL.26-30, THE POINTER  M4 WILL BE ZERO AND A MESSAGE IS   *
C     * PRINTED, INDICATING THE INABILITY TO SPLIT PARENT GROUP  K1     *
C     * ON PREDICTOR(I).IN THIS CASE THE NEXT PREDICTOR ENTER THE LOOP *
C     * ENDING WITH STATEMENT NO. 260                                   *
C     *
C     * IF AT LEAST ONE OF THE CLASSES QUALIFY, THE FOLLOWING ARE PRINTE
C     *               PRINT                       INTERNAL SYMBOL
C     *               MAX.BSS                     B
C     *               BSS/TSS                     S =B/GDATA(I,6)
C     *               BETWEEN CODES               ICODE(M),M=1,M4
C     *                AND    CODES               ICODE(M),M=L,L3
C     *
C     *   THE BSS IS COMPARED WITH PREVIOUS MAX.BSS  AND  THE LARGER-ON
```

```
C     *     IS STORED. NOW TRY THE NEXT PREDICTOR AND ITS CLASSES.
C     *
C     *
C     ****************************************************************
C
      IF (M3.LT.1) WRITE(19,9801) GOOGOL5,GOOGOL5
      IF (M3.LT.1) GO TO 210
      DO 200 K=1,M3
      J=ICODE(K)+M1
      N=SPLIT(J,1)
      NN=NN+N
      W1=SPLIT(J,2)+W1
      Y1=SPLIT(J,3)+Y1
      W2=W-W1
      Y2=Y-Y1
      A1=Y1/W1-AT
      A2=Y2/W2-AT
      S=Y1*A1+Y2*A2
      IF (L3.GT.KSTOP) GO TO 190
      PRINT 1590, ICODE(K),N,(SPLIT(J,L),L=2,4),AVER(K),SD(K)
      WSS = GDATA(1,6) - S
      PRINT 1600, S,WSS
  190 CONTINUE
      IF (NN.LT.NMIN) GO TO 200
      IF (NT-NN.LT.NMIN) GO TO 200
      IF (S.LE.B) GO TO 200
      B=S
      M4=K
  200 CONTINUE
      XXX=B/TTSS
      YAZ=B/GDATA(1,6)
      WRITE (19,9801) XXX,YAZ
  210 M3=M3+1
      IF (L3.GT.KSTOP) GO TO 220
      J=ICODE(M3)+M1
      N=SPLIT(J,1)
      PRINT 1590, ICODE(M3),N,(SPLIT(J,L),L=2,4),AVER(M3),SD(M3)
  220 S=B/GDATA(1,6)
      IF (M4.GT.0) GO TO 230
      IF (L3.GT.KSTOP) GO TO 260
      PRINT 1480, K1,I,(ACR(I,J),J=1,3)
      GO TO 260
  230 L=M4+1
      IF (L3.GT.KSTOP) GO TO 240
      PRINT 1460, B,S,(ICODE(M),M=1,M4)
      PRINT 1470, (ICODE(M),M=L,M3)
  240 CONTINUE
      IF (B.LT.BSS) GO TO 260
      BSS=B
      SS=S
      DO 250 K=1,M3
  250 INOUT(K)=ICODE(K)
C
C     ****************************************************************
C     *                                                              *
C     * INOUT(45) = PREDICTOR NUMBER WITH THE LARGEST BSS            *
C     * INOUT(44) = NUMBER OF CLASSES IN THIS PREDICTOR              *
C     *             WITH AT LEAST ONE CASE                           *
C     * INOUT(43) = CODE NUMBER WHERE SPLIT  OCCURS                  *
C     * INOUT(42) = CODE NUMBER OF THAT CLASS WITH THE LARGEST BSS   *
C     *                                                              *
C     ****************************************************************
C
      INOUT(45)=I
      INOUT(44)=M3
      INOUT(43)=L
      INOUT(42)=M4
  260 CONTINUE
      IDIOT=IDIOT+1
      I=INOUT(45)
      ITRIA(IDIOT)=K1
      IF (HSS.GE.RC) GO TO (350,480), IFSW2
      GO TO (270,280), IFSW2
  270 MAMA(K1,2)=1
  280 IF (I.EQ.0) GO TO 290
      IF(KSTOP.LT.0)GO TO 300
      PRINT 1490, K1,P2,I,(ACR(I,J),J=1,3),BSS,TTSS      ,RC
      GO TO 300
  290 K1=GDATA(1,1)
      IF(KSTOP.LT.0)GO TO 300
      PRINT 1500, K1
  300 IF (L2.EQ.0) GO TO 340
      F=GDATA(1,4)/GDATA(1,3)
      N=GDATA(1,2)
      S=SQRT(GDATA(1,6)/GDATA(1,3))
      K1=GDATA(1,1)
      IF(KSTOP.LT.0)GO TO 309
      PRINT 1420
      PRINT 1400, K1,N,(GDATA(1,K),K=3,6),F,S
  309 M5=K1
      DO 320 I=1,L2
      DO 310 J=1,6
  310 GDATA(I,J)=GDATA(I+1,J)
  320 CONTINUE
      L2=L2-1
      IA=3
      IB=3
      K4=GDATA(1,1)
      GO TO 880
  330 IF(KSTOP.LT.0)GO TO 339
      PRINT 1510, L3,L2
  339 IF ((RESID.EQ.0) GO TO 348
      IA=3
      IB=3
      KENDSW=2
      F=GDATA(1,4)/GDATA(1,3)
      M5=K1
      GO TO 880
  340 PRINT 1670
      IDF2=DF2
      PRINT 1680, IDF,TTSS,THSS,TWSS,TR2,TR,FX2,IDFF,IDF2
      IDUM=IPHASE+1
      RSUM(IDUM,1)=TR2
      RSUM(IDUM,2)=TR
      RSUM(IDUM,3)=FX2
      RSUM(IDUM,4)=IDFF
```

```
      MSUM(IDUM,5)=IDF2
      IF(ITREE.LT.1.OR.IFSW1.EQ.1)GO TO 349
      IL=1
      WRITE(19,9804) IDIOT
      K1=999999
      WRITE(10) ISTOR(5),SD(3),SD(7),IL,   INOUT(1),ISTOR(6),SD(4),SD(8)
     1,IL,   INOUT(1),IK,TR2,FX,T,K1
9800  FORMAT(E15.8)
9801  FORMAT(2E15.8)
9804  FORMAT(I3)
349   RETURN
350   K=0
      IFC=IFC+1
      IF(ICR(1,3).EQ.0) GO TO 359
      IFORCE(IFC,8)=I
      K=INOUT(42)
      IFORCE(IFC,2)=INOUT(K)
      K=INOUT(43)
      IFORCE(IFC,3)=INOUT(K)
      GO TO 480
359   NWF=INOUT(44)/IW
      IWF=INOUT(44)-NWF*IW
      IF (NWF.LT.1) GO TO 370
      DO 360 I=1,NWF
      IFORCE(IFC,I)=0
      DO 360 J=1,IW
      K=K+1
360   IFORCE(IFC,I)=40*IFORCE(IFC,I)+INOUT(K)
370   IF (IWF.EQ.0) GO TO 390
      I=NWF+1
      IFORCE(IFC,I)=0
      DO 380 J=1,IWF
      K=K+1
380   IFORCE(IFC,I)=40*IFORCE(IFC,I)+INOUT(K)
390   K=41
      IFORCE(IFC,8)=0
      DO 400 J=1,4
      K=K+1
400   IFORCE(IFC,8)=81*IFORCE(IFC,8)+INOUT(K)
      GO TO 480
410   IF(KSTOP.LT.0)GO TO 300
      PRINT 1640, K1
      GO TO 300
420   IF (MAMA(K1,2).EQ.1) GO TO 410
      IFC=IFC+1
      ITEMP=IFORCE(IFC,8)/81
      IK=IFORCE(IFC,8)-ITEMP*81
      IF(ICR(IK,3).EQ.0) GO TO 429
      M4=IFORCE(IFC,2)
      M3=IFORCE(IFC,3)
      M1=ICR(IK,2)
      M4=M1+M4
      M3=M1+M3
      L=ICR(IK,1)
      ISS1=1
      I=0
      DO 421 K=M1,M4
      IF(SPLIT(K,1).EQ.0.)GO TO 421
      I=I+1
      INOUT(I)=IABS(K - M1)
421   CONTINUE
      INOUT(42)=I
      IF(I.EQ.0) ISS1=2
      ISS2=1
      DO 422 K=M3,L
      IF(SPLIT(K,1).EQ.0.) GO TO 422
      I=I+1
      INOUT(I)=IABS(K - M1)
422   CONTINUE
      IF(I.EQ.INOUT(42)) ISS2=2
      INOUT(44)=I
      INOUT(45)=IK
      INOUT(43)=INOUT(42)+1
      GO TO 480
429   K=45
      DO 430 J=1,4
      ITEMP=IFORCE(IFC,8)/81
      INOUT(K)=IABS(IFORCE(IFC,8)-ITEMP*81)
      IFORCE(IFC,8)=ITEMP
430   K=K-1
      K=INOUT(44)
      NWF=K/IW
      IWF=K-NWF*IW
      IF (IWF.EQ.0) GO TO 450
      DO 440 J=1,IWF
      ITEMP=IFORCE(IFC,NWF+1)/40
      INOUT(K)=IABS(IFORCE(IFC,NWF+1)-ITEMP*40)
      IFORCE(IFC,NWF+1)=ITEMP
440   K=K-1
450   IF (NWF.LT.1) GO TO 470
      DO 460 I=1,NWF
      L=NWF-I+1
      DO 460 J=1,IW
      ITEMP=IFORCE(IFC,L)/40
      INOUT(K)=IABS(IFORCE(IFC,L)-ITEMP*40)
      IFORCE(IFC,L)=ITEMP
460   K=K-1
470   CONTINUE
      M1 = ICR(IK,2)
      M4 = INOUT(42)
      M3 = INOUT(44)
      MX = M3 - M4
      L = 1
      DO 473 K = 1,M4
      J = M1 + INOUT(L)
      IF (SPLIT(J,1) .GT. 0.) GO TO 472
      M3 = M3 - 1
      DO 471 J = L,M3
471   INOUT(J) = INOUT(J+1)
      GO TO 473
472   L = L+1
473   CONTINUE
      ISS1 = 1
      IF(L .EQ. 1) ISS1 = 2
      INOUT(42) = L-1
      DO 476 K = 1,MX
```

```
 M1 + INOUT(L)
SPLIT(J,1) .GT. 0.) GO TO 475
 M3 - 1
K3.EQ.0)GO TO 476
474 J = L,M3
T(J) = INOUT(J+1)
O 476
 L + 1
 INUE
 = 1
L .EQ. (.NOUT(42) + 1)) ISS2 = 2
T(44) = M3
T(43) = INOUT(42) + 1
T(45) = IK
NOUT(45)
CH(IK,2)
CR(IK,1)-M1+1
90 K=1,J
N(K)=999
NOUT(46)
3.EQ.0)GO TO 505
11 L=2,N4
OUT(K)+1
N(L)=K
IFS=1.EQ.2) GO TO 560
Y1S-GDATA(1,4)*GDATA(1,4)/GDATA(1,3)
510 K=2,5
)=0
530 I=1,J
INVIN(I).NE.999) GO TO 530
1+I-1
20 K=2,5
)=SD(K)+SPLIT(LL,K-1)
INUE
TNN-SD(2)
TNV-SD(3)
=GD14-SD(4)
=GD15-SD(5)
S=GD14*GD14/TNV
=GD15-TSYIS
40 K=2,5
A(1,K)=GDATA(1,K)-SD(K)
TSYIS+GDATA(1,4)*GDATA(1,4)/GDATA(1,3)
A(1,6)=GDATA(1,5)-GDATA(1,4)*GDATA(1,4)/GDATA(1,3)
0.
 = ACC1 - XM(K1)*SD(4)
 = ACC2 - XM(K1)*SD(3)
 = ACC3 - (XM(K1)**2)*SD(2)
SS1.EQ.2 .OR.ISS2.EQ.2)GO TO 585

ATA(1,3)
ATA(1,4)
/W
NOUT(42)
50 K=1,N4
OUT(K)+M1
1+SPLIT(J,2)
```

```
550   Y1=Y1+SPLIT(J,3)
      W2=W-W1
      Y2=Y-Y1
      BSS=W1*((Y1/W1-AT)**2)+W2*((Y2/W2-AT)**2)
      SS=BSS/GDATA(1,6)
      K2=2*L3
      K3=K2+1
      ACC1=ACC1+Y1*XM(K2)+Y2*XM(K3)-Y*XM(K1)
      ACC2=ACC2+W1*XM(K2)+W2*XM(K3)-W*XM(K1)
      ACC3=ACC3+W1*(XM(K2)**2)+W2*(XM(K3)**2)-W*(XM(K1)**2)
      ESSCR = GD15 - 2.*ACC1 + ACC3
      IF(IRSU1 .NE. 1)GO TO 555
      RRATV=1.-(GD15-2.*ACC1+ACC3)/(GD15-2.*GD14*XM(1)+TNV*(XM(1)**2))
555   IF(IRSU2 .NE. 1)GO TO 560
      RBVDOT=RBVDOT-BETA(K1)*(Y/W - GD14/TNV)*SQRT(W*TNV/((TNV-W)*TTSS))
     2+BETA(K2)*(Y1/W1-GD14/TNV)*SQRT(W1*TNV/((TNV-W1)*TTSS)) +
     3 BETA(K3)*(Y2/W2-GD14/TNV)*SQRT(W2*TNV/((TNV-W2)*TTSS))
C     ****************************************************************
560   L=INOUT(42)
      DO 580 I=1,3
      AVER(I)=0
      DO 570 J=1,L
      LL=INOUT(J)+M1
570   AVER(I)=AVER(I)+SPLIT(LL,I)
      AVER(I+3)=GDATA(1,I+1)-AVER(I)
580   CONTINUE
      A=(GDATA(1,6)-BSS)/GDATA(1,3)
      A=A*GDATA(1,2)/(GDATA(1,2)-2.0)
      B=1.0/AVER(1)+1.0/AVER(4)
      A=SQRT(A*B)
      B=AVER(3)/AVER(2)-AVER(6)/AVER(5)
      T=ABS(B/A)
      IF(GDATA(1,2) .LE. 2.) T = 0.
585   DO 600 I=1,L2
      DO 590 J=1,6
590   GDATA(I,J)=GDATA(I+1,J)
600   CONTINUE
      K=1
      K2=(2*L3)
      K3=K2+1
      L2=L2-1
      M2=1
      M3=1
      IF(KSTOP .GE. 0 .OR. LC .LE. 50) GO TO 629
      LC = 3
      PRINT 1710
629   IF (IFS=1.EQ.2) GO TO 610
      IF(KSTOP .GE. 0) PRINT 1540
      IF(ISS1.EQ.2.OR.ISS2.EQ.2)GO TO 630
      GO TO 620
630   PRINT 1689,K1,K2,K3
      IF(KSTOP .GE. 0) PRINT 1690,K2,K3
      LC = LC + 4
      GO TO 640
610   IF(KSTOP .GE. 0) PRINT 1535
620   PRINT 1530, K1,IK,(ACR(IK,J),J=1,3),K2,(INOUT(J),J=K,L)
      LL=INOUT(43)
      MM=INOUT(46)
```

```
 KSTOP .LT. 0) LC = LC+5+(L-1)/17 + (MM-LL  17
 NT 1520, K3,(INOUT(J),J=LL,MM)
 KSTOP .GE. 0) PRINT 1550,BSS,SS,T
 =IDF+1
 =IDF
 =INN-DFF
 S=TBSS+BSS+C
 =TBSS/TTSS
 SQRT(TR2)
 S=TTSS-TBSS
 (TR2-TR1)/((1.-TR2)/DF2)
 FX .LT. 0.) FX = 0.
 =(TBSS*DF2)/(TWSS*(DFF-1.))
 2=DF2+.5
 F=IDF-1
 IFSW1 .EQ. 2) GO TO 647
 = 0.
 ****************************************************************
 = (TNV*ACC1 - GD)4*ACC2)/SQRT(TNV*TTSS*(ACC3*TNV - ACC2**2))
 = TR*R
 = LC + LUNOPT
 IESS  .EQ. 1) PRINT 2005,ESSCR
 IRSQ1 .EQ. 1) PRINT 2006,RRATV
 IRSQ2 .EQ. 1) PRINT 2007,RBVDOT
                 KSTOP .LT. 0) PRINT 2008,TR
 MAT(79X,*ESS(FULL MODEL)*,10X,*=*,G17.8)
 MAT(79X,*1 - ESS(FULL)/ESS(REST.) =*,F12.8)
 MAT(79X,20HBETA(TRANS)*VALIDITY,5X,*=*F12.8)
 MAT(79X,*PROD. MOMENT CORRELATION =*,F12.8)
 ****************************************************************
 KSTOP.LT.0)GO TO 648
 NT 1660
 IFSW1.EQ.1)GO TO 691
 NT 1650, IDF,TTSS,TBSS,TWSS,TR2,TR,FX,IX1,IDF2,FX2,IDFF,IDF2
 TO 648
 NT 1651,IDF,TTSS,TBSS,TWSS,TR2,TR,FX2,IDFF,IDF2
 =TR2
 840 I= ,2
 650 J=1,4
 ER(J)=0
 ISS1.EQ.1.AND.ISS2.EQ.1)GO TO 659

 TO (652,654),I
 ISS1.EQ.2) GO TO 671
 TO 658
 ISS2.EQ.2) GO TO 671
 NOUT(44)
 670 J=K,L
 =INOUT(J)+M1
 660 M=1,4
 ER(M)=AVER(M)+SPLIT(LL,M)
 TINUE
 TINUE
 VER(3)/AVER(2)
 IFSW2 .EQ. 1) XM(K2) = A
 IFSW2 .EQ. 1) BETA(K2) = A*SQRT(AVER(2)*(TNV-AVER(2))/
```

```
     1(TNV*TTSS))
C    ****************************************************************
     S=AVER(4)-A*AVER(3)
671  F2=F3
     SD(I+2)=A
     ISTOR(I+4)=AVER(1)+.5
     SD(I+6)=SQRT(S/(AVER(2)+1.E-9))
     M=M2
     M2=M3
     M3=M
     N=AVER(1)/2.0
     GO TO (680,690), IFSW1
680  IF (MAMA(K2,1).EQ.999) GO TO 790
     GO TO 700
690  IF (N.LT.NMIN) GO TO 790
     IF (S.LT.RE.OR. S .EQ. 0.) GO TO 790
700  IF (L2.EQ.0) GO TO 760
     DO 750 J=1,L2
     M=L2-J+1
     GO TO (710,720), IFSW1
710  IF (M.LT.MAMA(K2,1)) GO TO 770
     GO TO 730
720  IF (S.LE.GDATA(M,6)) GO TO 770
730  DO 740 MM=1,6
     GDATA(M+1,MM)=GDATA(M,MM)
740  CONTINUE
750  CONTINUE
760  M=0
770  M=M+1
     IF (IFSW2.EQ.1) MAMA(K2,1)=M
     GDATA(M,1)=K2
     DO 780 MM=2,5
     GDATA(M,MM)=AVER(MM-1)
780  CONTINUE
     GDATA(M,6)=S
     GO TO 830
790  M3=2
     IF (IFSW2.EQ.1) MAMA(K2,1)=999
     IF (IRESID.EQ.0) M3=3
     F3=A
     IF (KSTOP.LT.0)GO TO 829
     IF (IFSW1.EQ.1) GO TO 810
     IF (N.LT.NMIN) GO TO 800
     PRINT 1560, K2,P1,S,TTSS,RE
     GO TO 820
800  PRINT 1570, K2,AVER(2),NMIN
     GO TO 820
810  PRINT 1640, K2
820  PRINT 1420
     G=AVER(1)
     B=SQRT(S/AVER(2))
     IF(AVER(2).EQ.0.)B=0.
     PRINT 1400, K2,N,(AVER(M),M=2,4),S,A,B
829  L2=L2-1
830  L2=L2+1
     K2=K2+1
     K=INOUT(43)
     L=INOUT(44)
```

```
      SU(I)=S
      L3=L3+1
      K2=K2-2
      IF(KSTOP.LT.0)PRINT 1700,K2,ISTOR(5),SD(3),K3,ISTOR(6),SD(4),T,
     1TR2,FX
      IM=INOUT(42)
      IL=1
      NC1=IM-IL+1
      NC2=L-K+1
      IF(ITREE.GE.1.AND.IFSW2.ME.1.AND.K3.LE.179)WRITE(10) ISTOR(5),SD(3
     1),SD(7),NC1  ,(INOUT(J),J=IL,IM),ISTOR(6),SD(4),SD(8),NC2,(INOUT(J
     2),J=K,L),IK,TR2,FX,T,K1
      ILL=1
      K4=GDATA(1,1)
      IF (SD(1).GE.SD(2)) GO TO 860
      IF (K4.NE.K3) GO TO 870
      A=F2
      F2=F1
      F3=A
      M=M3
      M3=M2
      M2=M
      K2=K1
      K3=K1-1
      IM=INOUT(44)
      IL=INOUT(43)
      ILL=2
      GO TO 870
      CONTINUE
      IF (K4.NE.K2) GO TO 850
      IP=IK/IW
      IF (IW*IP.LT.IK) IP=IP-1
      IQ=IK-IW*IP+6
      IQ=IOCT(IQ)
      IF (IRESID.GT.0) IP=IP+1
      K4=GDATA(1,1)
      IA=1
      IB=1
      IF (K4.EQ.K2) GO TO 890
      IB=2
      L=999999
      WRITE (KD) L,(ISTOR(J),J=1,KBL)
      REWIND KD
      L=KO
      KO=KH
      KH=L
      GO TO (890,940), KENDSW
      IF (L2.EQ.0) GO TO 915
      IF(KSTOP .LT. 0) GO TO 918
      PRINT 1410
      PRINT 1420
      DO 900 J=1,L2
      I=GDATA(J,1)
      N=GDATA(J,2)
      A=GDATA(J,4)/GDATA(J,3)
      B=SQRT(GDATA(J,6)/GDATA(J,3))
      PRINT 1400, I,N,(GDATA(J,K),K=3,6),A,B
      DO 930 J=1,700
      DO 920 K=1,4
920   SPLIT(J,K)=0
930   CONTINUE
940   JOE2=0
      IP2=51
      L=KA
      KA=KC
      KC=L
      LV=156
      LW=311
      IC=1
950   KV=NBL+155
      KW=NBL+310
960   READ (KA) L,(ISTOR(J),J=1,KBL)
      LM=NBL
      DO 1270 LX=1,JBL,2
      IF (L.NE.K1) GO TO 990
970   IF (ISTOR(LM).EQ.999999) GO TO (1120,1280,960), IA
      W=ISTOR(LM)
      LM=LM+1
      DO 980 J=1,NL
      ICODE(J)=ISTOR(LM)
      LM=LM+1
980   CONTINUE
      T=AVER(LX)
      TT=AVER(LX+1)
      GO TO (1040,1250,1200), IA
990   IF (L.GT.900) GO TO 1120
      GO TO (1000,1010), IC
1000  WRITE (KD) L,(ISTOR(J),J=1,KBL)
      GO TO 960
1010  IF (IRESID.EQ.0) GO TO 1300
      K1=L
      DO 1020 J=1,L3
      K=GDATA(J,1)
      IF (K.EQ.L) GO TO 1030
1020  CONTINUE
      GO TO 340
1030  F=GDATA(J,4)/GDATA(J,3)
      MS=-1
      GO TO 970
1040  K=ICODE(IP)/IQ
      LL=K/40
      K=K-40*LL
      IF (INVIN(K+1).NE.999) GO TO (1060,1070), ILL
      IF(KSTOP .LT. 0) GO TO 1270
      ID=ICODE(1)
      IF (IRESID.EQ.0) ID=0
      K=IABS(K)
      IF (IP2.LE.50) GO TO 1050
      IP2=0
      PRINT 1410
1050  PRINT 1500, ID,K1,K
      IP2=IP2+1
      GO TO 1270
1060  IF (INVIN(K+1).LE.IM) GO TO (1160,1140,1270), M2
      GO TO 1080
1070  IF (INVIN(K+1).GE.IL) GO TO (1160,1140,1270), M2
```

```
      GO TO (1090,1150,1270), M3
      AVER(LW)=Y
      AVER(LW+1)=YY
      ISTOR(KW)=W
      KW=KW+1
      DO 1100 J=1,NL
      ISTOR(KW)=ICODE(J)
      KW=KW+1
      CONTINUE
      LW=LW+2
      IF (KW.LT.MBL) GO TO 1270
      LW=311
      WRITE (KC) K3,(ISTOR(J),J=311,MBL)
      KW=MBL+310
      GO TO (1270,1130), IC
      IC=2
      IF (LW.EQ.311) GO TO 1130
      ISTOR(KW)=999999
      GO TO 1110
      IF (LV.EQ.156) GO TO 1300
      ISTOR(KV)=999999
      GO TO 1190
      MS=K2
      XLL=Y/W-F2
      LL=XLL+SIGN(.5,XLL)
      GO TO 1210
      MS=K3
      XLL=Y/W-F3
      LL=XLL+SIGN(.5,XLL)
      GO TO 1210
      AVER(LV)=Y
      AVER(LV+1)=YY
      ISTOR(KV)=W
      KV=KV+1
      DO 1170 J=1,NL
      ISTOR(KV)=ICODE(J)
      KV=KV+1
      CONTINUE
      LV=LV+2
      IF (KV.LT.LBL) GO TO 1190
      WRITE (KC) K2,(ISTOR(J),J=156,LBL)
      KV=NBL+155
      LV=156
      GO TO (1190,1300), IC
      GO TO (1250,1270,960), IB
      IF (IRESID.EQ.0) GO TO 960
      XLL=Y/W-F
      LL=XLL+SIGN(.5,XLL)
      M=W
      GO TO (1220,1240,1240), IRESID
      ITIME=1
      JOE2=JOE2+1
      J22(JOE2,1)=MS
      J22(JOE2,2)=ICODE(1)
      J22(JOE2,3)=M
      J22(JOE2,4)=LL
      IF (JOE2.LT.4) GO TO 1270
      IF (IP2.LE.50) GO TO 1230
      IP2=0
      PRINT 1610
1230  PRINT 1620, ((J22(II,KK),KK=1,4),II=1,4)
      IP2=IP2+1
      JOE2=0
      GO TO 1270
1240  WRITE (KF,1630) MS,ICODE(1),M,LL
      GO TO 1270
1250  I=0
      DO 1260 J=N1,NL
      LL=ICODE(J)
      DO 1260 K=IV,6
      IF (I.EQ.NP) GO TO 1270
      M=LL/IOCT(K)
      LL=LL-M*IOCT(K)
      I=I+1
      KK=ICH(I,2)+M
      SPLIT(KK,1)=SPLIT(KK,1)+1.0
      SPLIT(KK,2)=SPLIT(KK,2)+W
      SPLIT(KK,3)=SPLIT(KK,3)+Y
      SPLIT(KK,4)=SPLIT(KK,4)+YY
1260  CONTINUE
1270  CONTINUE
      GO TO (960,1280,960), IA
1280  GO TO (960,1290), IFSW
1290  WRITE (KC) L,(ISTOR(J),J=1,KBL)
      GO TO 960
1300  REWIND KA
      IF (ITIME.NE.1) GO TO 1310
      IF (IP2.GT.50) PRINT 1610
      IF (JOE2.NE.0) PRINT 1620, ((J22(II,KK),KK=1,4),II=1,JOE2)
      ITIME=2
1310  CONTINUE
      K1=K4
      L=999999
      WRITE (KC) L,(ISTOR(J),J=1,KBL)
      REWIND KC
      GO TO (1360,1320,1340), IB
1320  READ (KC) L,(ISTOR(J),J=1,KBL)
      IF (L.GT.900) GO TO 1330
      WRITE (KD) L,(ISTOR(J),J=1,KBL)
      GO TO 1320
1330  REWIND KC
1340  CONTINUE
      IF (L2.EQ.0) GO TO 340
      LV=156
      LW=311
      L=KA
      KA=KB
      KB=L
      IF (KENDSW.EQ.2) GO TO 1350
      IC=1
      IB=1
      IA=2
      GO TO 950
1350  IC=2
      IA=3
      IB=3
```

```
L3=L2
L2=0
F1=0
GO TO 950
GO TO (1370,55), IFS=
FS==2
FS=1=1
I=ICR(2,2)-1
DO 1380 I=2,5
GDATA(1,I)=0
DO 1390 I=1,K
GDATA(1,2)=GDATA(1,2)+SPLIT(I,1)
GDATA(1,3)=GDATA(1,3)+SPLIT(I,2)
GDATA(1,4)=GDATA(1,4)+SPLIT(I,3)
GDATA(1,5)=GDATA(1,5)+SPLIT(I,4)
GDATA(1,1)=1
GDATA(1,6)=GDATA(1,5)-(GDATA(1,4)*GDATA(1,4))/GDATA(1,3)
PRINT 1420
I=GDATA(1,1)
N=GDATA(1,2)
A=GDATA(1,4)/GDATA(1,3)
B=SQRT(GDATA(1,6)/GDATA(1,3))
PRINT 1400, I,N,(GDATA(1,K),K=3,6),A,B
L=KA
KA=KC
KC=L
REWIND KA
REWIND KB
REWIND KC
REWIND KD
ACC1 = XM(1)*GDATA(1,4)
ACC2 = XM(1)*GDATA(1,3)
ACC3 = (XM(1)**2)*GDATA(1,3)
IXI = 0
GO TO 40

FORMAT (2I7,6G17.8)
FORMAT (/36H   CANDIDATE GROUPS ARE  AS FOLLOWS.)
FORMAT (/8H   GROUP,4X,1HN,5X,12HTOTAL WEIGHT,8X,8HSUM OF Y,6X,12H
SUM Y-SQUARE,10X,5HT S S,12X,4HMEAN,10X,9HSTD. DEV.)
FORMAT (8H    CODE,4X,1HN,5X,12HTOTAL WEIGHT,8X,8HSUM OF Y,6X,12HS
UM Y-SQUARE,10X,5HMEAN ,12X,9HSTD. DEV., 9X,5HB S S,10X,5HW S S)
FORMAT (13H1** STEP NO.=,I3,7X,14HPARENT GROUP =,I3,3H **)
FORMAT (18H  TRY ON PREDICTOR,I3,3X,3A6)
FORMAT (10H0MAX. BSS=,G18.8,5X,9HBSS/TSS =,F10.5,3X,14HBETWEEN COD
.ES ,20I3/69X,20I3)
FORMAT (55X,10HAND CODES ,4X,20I3/69X,20I3)
FORMAT (1H0,*GROUP*,I4,2X,*CANNOT BE SPLIT ON PREDICTOR*,I4,3X,3A6
1,5X,*FOR FAILURE TO MEET MINIMUM NUMBER OF OBSERVATIONS REQUIREMEN
2TS*)
FORMAT (1H1,131(1H*)/*0GROUP*,I3,* CANNOT BE SPLIT FOR FAILURE TO
1REDUCE THE UNEXPLAINED SUM OF SQUARES BY P2=*,F9.6,5X,/1H ,*PROPOR
2TION OF THE TOTAL SUM OF SQUARES SPECIFIED BY THE USER ON CARD 3,
3COL. 7-11*,/1H ,*MAX.BSS.FOR THIS GROUP ON PREDICTOR*,I3,2X,3A6,6X
4,*=*,E20.8,/1H ,60X,*TSS =*,E20.8,/1H ,57X,8HP2*TSS =,E20.8,/1H ,1
531(1H*)//)
FORMAT (1H1,131(1H*)/1H0,*ALL PREDICTORS FOR GROUP*,I3,5X,*FAILED
1TO MEET THE MIN. NUMBER OF OBSERVATION REQUIREMENTS*/1H0,131(1H*,)

1510 FORMAT (*1EITHER NUMBER OF FINAL GROUPS EQUALS MAXGP OR NUMBER OF
    1CANDIDATE GROUPS EQUALS 90.TERMINATE RUN*//* NUMBER OF FINAL GROUP
    2S = *,I3,* NUMBER OF CANDIDATE GROUPS = *,I3)
1520 FORMAT (55X,9HAND GROUP,I5,11H WITH CODES,17I3/79X,17I3/79X,17I3)
1530 FORMAT (                1H ,*SPLIT GROUP *I3,* ON PREDICTOR*,I4,3X,3A
    16,* INTO GROUP*,I4,* WITH CODES*,17I3/79X,17I3/79X,17I3)
1535 FORMAT(1H1,131(1H*))
1540 FORMAT(1H0,131(1H*))
1550 FORMAT (/9X,6HBSS  =,G18.8,13H   BSS/TSS  =,F10.5,10H   T-VALUE,F6
    1.2)
1560 FORMAT (1H0,131(1H*)/*0GROUP*,I3,X,*CANNOT BE SPLIT FOR FAILURE TO
    1 CONTAIN P1=*F9.6,5X,*PROPORTION OF THE TOTAL SUM OF SQUARES*/* AS
    2 SPECIFIED BY THE USER ON CARD 3, COL.  2-6.                     TS
    3S IN THIS GROUP=*,E20.8,/1H ,80X,*TSS=*,E20.8,/1H ,76X,8HP1*TSS =,
    4E20.8/1H0,131(1H*)//)
1570 FORMAT (1H0,131(1H*)/1H0,*GROUP*,I3,X,*CANNOT BE SPLIT.THE NUMBER
    1OF CASES IN ANY  SPLIT IS LESS THAN THE SPECIFIED MINIMUM*/1H0,*NU
    2MBER OF CASES IN THIS GROUP=*,F8.0/1H ,*MINIMUM SPECIFIED BY USER
    3   = *,I6,2',*FOR ANY GIVEN SPLIT*/1H0,131(1H*)//)
1580 FORMAT (6H *CASE,I10,* IN GROUP *,I3,* ELIMINATED WITH RESPONSE CO
    1DE *,I3)
1590 FORMAT (2I7,5G17.8)
1600 FORMAT (4X,95(1H*),2G16.8)
1610 FORMAT (1H-,50X,*RESIDUAL AND/OR REJECT SUMMARY*,/1H0,4(*GROUP
    11S    WT       RES         *)/X,131(1H*))
1620 FORMAT (4(I3,I9,I6,I9,6X))
1630 FORMAT (I3,I9,I6,I9,53X)
1640 FORMAT (//*)*,131(1H*)//* GROUP *,I3,* WAS NOT SPLIT *//*0*,131(1H
    1*)//)
1650 FORMAT (//* NCF       TOTAL TSS          TOTAL BSS          TOTAL WSS
    1    R-SQUARED          R            F-RSQ      DF1    DF2      F-ANOVA    DF1
    2 (F2*//1X,I3,3G17.8,F13.8,F10.5,F12.4,I5,I7,F12.4,I5,I7,//*0*,131(
    31H*))
1651 FORMAT (//* NCF       TOTAL TSS          TOTAL BSS          TOTAL WSS
    1    R-SQUARED          R                                     F-ANOVA    DF1
    2 (F2*//1X,I3,3G17.8,F13.8,F10.5,24X          ,F12.4,I5,I7,//*0*,131(
    31H*))
1660 FORMAT (*0*,131(1H*)//58X,*CURRENT SUMMARY*//)
1670 FORMAT (*1*,131(1H*)//58X,*FINAL SUMMARY*//)
1680 FORMAT (//* NCF       TOTAL TSS          TOTAL BSS          TOTAL WSS
    1    R-SQUARED          R                                     F-ANOVA    DF1
    2 (F2*//1X,I3,3G17.8,F13.8,F10.5,24X          ,F12.4,I5,I7,//*0*,131(
    31H*))
1689 FORMAT(                   * SPLIT GROUP*,I4,* INTO GROUP*,I4,
    2* AND GROUP*,I4)
1690 FORMAT (*0BSS IS UNDEFINED SINCE EITHER GROUP*,I4,* OR GROUP*,I4,*
    1 CONTAINS NO CASES. NOTE THAT THE NUMBER OF FINAL GROUPS IS NOT IN
    2CREASED.*)
1700 FORMAT(7X,*GROUP *,I3,*  N=*,I6,* MEAN=*,F9.2,8X,*GROUP *,I3,
    1*  N=*,I6,* MEAN=*,F9.2,4X,*T=*,G11.4,* RSQ=*,F8.5,
    2* F-RSQ=*,G11.4/1H ,131(1H*))
1710 FORMAT(1H1,59X,*SPLIT SUMMARY*/)
1711 FORMAT(* GROUP 1 DID NOT SPLIT.TERMINATE*)
     END
```

```
      SUBROUTINE AIDFIN
      DIMENSION BSS(1950)
      DIMENSION NAMI(30),SUM(30),MEANBSS(30),COEFF(30)
      DIMENSION SD(30),VARIANC(30),RELIMP(30),RANK(30),GROUPS(30)
      DIMENSION OLD(30),LABELS(7)
      DIMENSION TT(2)
      COMMON /APT/NP,NAMI,ITRIA(163)
      TYPE REAL MEANBSS
      TYPE REAL NAMI
      DATA (SUM=30(0.0)),(SD=30(0.0))
      DATA ROUNDER /.01/
      DATA(LABELS = 6HSUBSUM,6HTRIALS,6HMEANBS,6H   S D,6HCF VAR,
     66H   VAR,6HRECSUM)
      GBSMEAN =0.
      GOOGOLS=10.**30
      BACKSPACE 19
      TT(1)=10HBSS/TSST
      TT(2)=10HBSS/TSS(I)
      ICHEK =0
      NV=NP
      CONTINUE
      DO 1024 I=1,NV
      SUM(I)=0.
      SD(I)=0.
      CONTINUE
      GBSMEAN=0.
      READ(19,10)NGROUP  $  REWIND 19

      READ(19,13)TSST
      KONST=NV*NGROUP
      KO=ICHEK+1
      IF(ICHEK)1,1,2
      DO 31 I=1,KONST
      READ(19,24)BSS(I),XX
      CONTINUE
      GO TO 
      DO 32 I=1,KONST
      READ( ,24)XX,BSS(I)
      CONTINUE
      CONTINUE
      DO 1015 I=1,NV
      GROUPS(I)=NGROUP
      DO 1013 J=1,NGROUP
      INDEX=(J-1)*NV+I
      ABC=BSS(INDEX)
      IF(ABC.GE.GOOGOLS) GO TO 10125
      SUM(I)=SUM(I)+ABC
      SD(I)=SD(I)+ABC*ABC
      GO TO 1013
      GROUPS(I)=GROUPS(I)-1
      BSS(INDEX)=0.
      CONTINUE
      CONTINUE
      DO 102 I=1,NV
      MEANBSS(I)=SUM(I)/GROUPS(I)
      GBSMEAN =GBSMEAN +MEANBSS(I)
      VARIANC(I)=(SD(I)-MEANBSS(I)*MEANBSS(I)*GROUPS(I))/GROUPS(I)
      SD(I)=SQRT(VARIANC(I))
  102 CONTINUE
      GBSMEAN=GBSMEAN/NV
      DO 103 I=1,NV
      RELIMP(I)=MEANBSS(I)/GBSMEAN *100
      COEFF(I)=SD(I)/MEANBSS(I)
  103 CONTINUE
      GN=NGROUP
      DO 107 I=1,NV
      OLD(I)=MEANBSS(I)
  107 MEANBSS(I)=MEANBSS(I)*GN
      GBSMEAN=GBSMEAN*GN
      NPG=1+(NV-1)/6
      DO 101 N=1,NPG
      J1=1+(N-1)*6    $ J2 = MIN0(J1+5,NV)
      PRINT 27,TT(KO)
      PRINT 15,(NME,(NAMI(NME)),NME=J1,J2)
      DO 100 J=1,NGROUP
      NIX=(J-1)*NV
      N1=NIX+J1
      N2=NIX+J2
      KJ=ITRIA(J)
      PRINT 16, J,KJ,(BSS(I),I=N1,N2)
  100 CONTINUE
      PRINT 20
      PRINT 20,LABELS(1),(SUM(I),I=J1,J2)
      PRINT 22,LABELS(2),(GROUPS(I),I=J1,J2)
      PRINT 20, LABELS(3),(OLD(I),I=J1,J2)
      PRINT 20, LABELS(4),(SD(I),I=J1,J2)
      PRINT 20,LABELS(5),(COEFF(I),I=J1,J2)
      PRINT 20,LABELS(6),(VARIANC(I),I=J1,J2)
      PRINT 20,LABELS(7),(MEANBSS(I),I=J1,J2)
  101 CONTINUE
      PRINT 17    ,NV,NGROUP,GBSMEAN,ROUNDER,TSST,TT(KO),TT(KO)
      CALL RANKSIT(RELIMP,RANK,NV,ROUNDER)
      DO 104 I=1,NV
      PRINT 19,    I,NAMI(I),SUM(I),MEANBSS(I),SD(I),COEFF(
     1I),VARIANC(I),RELIMP(I),GROUPS(I),RANK(I)
  104 CONTINUE
      IF(ICHEK)51,51,52
   51 ICHEK=1
      GO TO 50
   52 CONTINUE
10    FORMAT(I3)
   11 FORMAT(I3,F5.2,I1)
   13 FORMAT(F15.8)
C  THIS MUST ALSO BE THE FORMAT USESD IN THE OUTPUT FROM AID
C
   14 FORMAT(*1*,10A8)
   15 FORMAT(*0*,10X,6(I2,A10,7X)///* TRIAL/GRP*)
   16 FORMAT(I4,3X,I3,6(F13.6,6X))
   27  FORMAT(*1*25X,*ANALYSIS WITH    *A10)
   17 FORMAT(*1SUMMARY TABLE *,I4,* VARIABLES, NO OF SUBGROUPS IS *
     S    ,I4,5X,*GRAND MEAN = *F11.6,* ROUND CRIT. =*F6.3,*TSST =*F8.1
     1///25X,*SUB(VARIANT)  RECONSTRUCTED*/25X,*SUM*,A10,*  SUM*,A10,
     2* S D    COEFF VAR  VARIANCE PCT GRD MEAN  N GROUPS      RANK*;

   18 FORMAT(I6,3X,A10,7X,6F11.5,F11,4X,F11.1)
```

```
FORMAT(5X,I2,F12.5,F10.F10.1)
FORMAT(* *,A6,F15.5,5(F19.5))
FORMAT(* *,A6,F15,4X,5(F15,4X))
FORMAT(I6,3X,A10,2X,F15.5,F15,F15.1)
FORMAT(2E15.8)
RETURN
END
```

```
SUBROUTINE RANKSIT(SET,RANK,N,ROUNDER)
DIMENSION SET(30),RANK(30),ORD(100),X(100)
INTEGER ORD,ORDSUBI,ORDSUBJ
DO 1 I=1,N
X(I)=SET(I)
ORD(I)=I
CONTINUE
X(N+1)= -1.
DO 2 I=1,N
DO 2 J=I,N
IF(X(I).GT.X(J)) GO TO 2
STORE=X(I)   $X(I)=X(J)   $X(J)=STORE
STORE =ORD(I)   $  ORD(I)=ORD(J)   $ ORD(J)=STORE
CONTINUE
I=0
I=I+1
IF(I.GT. N) GO TO 6
ORDSUBI=ORD(I)
RANK(ORDSUBI)=I

ICOUNT =1
J=I
CONTINUE
IF X(J)-ROUNDER.GT.X(J+1)) GO TO 4
J=J+1
ICOUNT =ICOUNT+1
GO TO 3
CONTINUE
J=I-1
DO 5 K=1,ICOUNT
ORDSUBI=ORD(J+K)
RANK(ORDSUBI)=I (ICOUNT-1)/2
CONTINUE
I=J+ICOUNT
GO TO 7
CONTINUE
RETURN
 END
```

AIDTRE

```
      PROGRAM TREEC(INPUT,OUTPUT,TAPE10,TAPE14,TAPE19)
      COMMON IRUN,ITREE,NG1,XMG1,SDG1,NG2,XMG2,SDG2,ACR(80,3),NOGO
C     MODS TO TREEC
      DIMENSION HI(183),ID(183)
      INTEGER HI
      DIMENSION N(181),XMEAN(181),SD(181),ICODE(181,39),IPRED(180),
     1RSQ(180),FRSQ(180),T(180),KIDS(179,2),
     2NCODE(179),IWORK(2),IWORK1(4),IWORK2(8),IWORK3(64)
      COMMON/ELT/N,XMEAN,SD,ICODE,IPRED,RSQ,FRSQ,T,NCODE,KIDS
      REWIND 14
      READ(14) IRUN,ITREE,NG1,XMG1,SDG1,NG2,XMG2,SDG2,ACR,NOGO
      REWIND 10
      IFH=1
      N(1)=NG1
      XMEAN(1)=XMG1
      SD(1)=SDG1
3001  FORMAT(*1START OF TREE FOR SAMPLE A*)
1000  DO 1001 I=1,179
      DO 1001 J=1,2
1001  KIDS(I,J)=0
      DO 1010 I=2,183,2
      READ(10) N(I),XMEAN(I),SD(I),NC1,(ICODE(I,J),J=1,NC1),N(I+1),
     1XMEAN(I+1),SD(I+1),NC2,(ICODE(I+1,J),J=1,NC2),IPRED(I),RSQ(I),
     2FRSQ(I),T(I),K1
      IF(N1.EQ.999999)GO TO 1020
      HI(K1)=I+1
      IF(XMEAN(I).GT.XMEAN(I+1)) HI(K1)=I
      ID(I)=K1    $ID(I+1)=K1     $NOGPT=(I+1)
      NCODE(I)=NC1
      NCODE(I+1)=NC2
      KIDS(K1,1)=I
      KIDS(K1,2)=I+1
      IPRED(I+1)=IPRED(I)
      RSQ(I+1)=RSQ(I)
      FRSQ(I+1)=FRSQ(I)
1010  T(I+1)=T(I)
1020  IPHASE=1
      IF(ITREE.GE.2) CALL AIDPLT(HI,ID,NOGPT)
      IF(ITREE.EQ.3) CALL EXIT
      IF(ITFR.EQ.1)  PRINT 3001
      IF(ITFR.EQ.2)  PRINT 3000
      IFH=1
      DO 1021 I=1,K
1021  IWORK2(I)=0
      GO TO 1050
1030  IPHASE=2
      DO 1031 I=1,64
1031  IWORK3(I)=0
      J=0
1040  J=J+1
      IF(J.GT.K)GO TO 1045
      IGP=IWORK2(J)
      IF(IGP.EQ.0)GO TO 1040
      IF(KIDS(IGP,1).EQ.0)GO TO 1040
      L=(J-1)*8
      GO TO 1050
1045  IPHASE=3
      J=0
```

```
1046  J=J+1
      IF(J.GT.64)GO TO 2000
      IGP=IWORK3(J)
      IF(IGP.EQ.0)GO TO 1046
      IF(KIDS(IGP,1).EQ.0)GO TO 1046
      DO 1047 I=1,8
1047  IWORK2(I)=0
1050  CALL PLOTB
      CALL PLOTV(21,19,21,21H*******        *******)
      CALL PLOTV(3,25,9,9H******** )
      CALL PLOTH(4,25,35,35H***********************************)
      CALL PLOTH(4,33,35,35H***********************************)
      CALL PLOTV(39,25,9,9H*********)
      CALL PLOTHI(4,28,6,6HGROUP ,IGP,3)
      CALL PLOTHF(20,28,5,5HMEAN=,XMEAN(IGP),9,2)
      CALL PLOTHI(5,29,2,2HN=,N(IGP),6)
      CALL PLOTHF(20,29,5,5HS.D.=,SD(IGP),9,2)
      CALL PLOTHI(13,30,5,5HLEVEL,IPHASE,2)
      IWORK(1)=0
      IWORK(2)=0
      DO 1051 I=1,4
1051  IWORK1(I)=0
      NUM1=KIDS(IGP,1)
      NUM2=KIDS(IGP,2)
      IF(XMEAN(NUM1).GT.XMEAN(NUM2))GO TO 50
      IWORK(1)=NUM2
      IWORK(2)=NUM1
      GO TO 60
50    IWORK(1)=NUM1
      IWORK(2)=NUM2
60    DO 70 I=1,2
      IGP=IWORK(I)
      CALL BLOCK(I+1,IGP)
      NUM1=KIDS(IGP,1)
      IF(NUM1.EQ.0)GO TO 70
      NUM2=KIDS(IGP,2)
      IF(XMEAN(NUM1).GT.XMEAN(NUM2))GO TO 65
      IWORK1(2*I-1)=NUM2
      IWORK1(2*I)=NUM1
      GO TO 70
65    IWORK1(2*I-1)=NUM1
      IWORK1(2*I)=NUM2
70    CONTINUE
      DO 80 I=1,4
      IGP=IWORK1(I)
      IF(IGP.EQ.0)GO TO 80
      CALL BLOCK(I+3,IGP)
      NUM1=KIDS(IGP,1)
      IF(NUM1.EQ.0)GO TO 80
      NUM2=KIDS(IGP,2)
      IF(XMEAN(NUM1).GT.XMEAN(NUM2))GO TO 75
      IF(IPHASE.EQ.2)GO TO 71
      IWORK2(2*I-1)=NUM2
      IWORK2(2*I)=NUM1
      GO TO 80
71    K=2*I+L
      IWORK3(K-1)=NUM2
      IWORK3(K)=NUM1
      GO TO 80
75    IF(IPHASE.EQ.2)GO TO 76
      IWORK2(2*I-1)=NUM1
      IWORK2(2*I)=NUM2
      GO TO 80
76    K=2*I+L
      IWORK3(K-1)=NUM1
      IWORK3(K)=NUM2
80    CONTINUE
      DO 90 I=1,8
      IF(IPHASE.EQ.2)GO TO 81
      IGP=IWORK2(I)
      GO TO 82
81    K=L+I
      IGP=IWORK3(K)
82    IF(IGP.EQ.0)GO TO 90
      CALL BLOCK(I+7,IGP)
90    CONTINUE
      CALL PLOTP
      GO TO (1030,1040,1046),IPHASE
2000  IF(IRUN.NE.3)GO TO 2010
      IF(ITER.EQ.2)GO TO 2010
      ITER=ITER+1
      N(1)=NG2
      XMEAN(1)=XMG2
      SD(1)=SDG2
3000  FORMAT(*1START OF TREE FOR SAMPLE 8*)
      GO TO 100C
2010  CALL EXIT
      END
```

```
      SUBROUTINE BLOCK(NBLOCK,IGROUP)
      COMMON IRUN,ITREE,NG1,XMG1,SDG1,NG2,XMG2,SDG2,ACR(80,3),NOGO
      COMMON/ELT/N,XMEAN,SD,ICODE,IPRED,RSQ,FRSQ,T,NCODE,KIDS
      DIMENSION N(181),XMEAN(181),SD(181),ICODE(181,39),IPRED(180),
     1RSQ(180),FRSQ(180),T(180),KIDS(179,2),
     2NCODE(179),IX(15),IY(15)
      DATA KCOM/0/,IX/3,3,3,46,46,46,46,89,89,89,89,89,89,89,89/,IY/25,
     111,45,4,18,33,47,1,8,15,22,29,36,43,50/
      IRTPOS=IX(NBLOCK)
      LINE=IY(NBLOCK)
      CALL PLOTV(IRTPOS,LINE,8,8H********)
      CALL PLOTH(IRTPOS,LINE,44,44H************************************
     1********)
      CALL PLOTV(IRTPOS+43,LINE,8,8H********)
      CALL PLOTH(IRTPOS,LINE+7,44,44H***********************************
     1*********)
      CALL PLOTHI(IRTPOS+1,LINE+1,6,6HGROUP ,IGROUP,3)
      IF(KIDS(IGROUP,1).EQ.0)CALL PLOTH(IRTPOS+10,LINE+1,6,6H*FINAL)
      CALL PLOTHF(IRTPOS+17,LINE+1,5,5HMEAN=,XMEAN(IGROUP),9,2)
      CALL PLOTHF(IRTPOS+32,LINE+1,5,5HRSQ = RSQ(IGROUP),6,3)
      CALL PLOTHI(IRTPOS+ 2,LINE+2,2,2HN=,N(IGROUP),6)
      CALL PLOTHF(IRTPOS+17,LINE+2,5,5HS.D.=,SD(IGROUP),9,2)
      IF(KCOM.EQ.1)GO TO 1000
      DF2=N(1)-(IGROUP/2+1)
      IF(DF2.GT.1000.)DF2=1000.
      CALL PLEVEL(1.,DF2,FRSQ(IGROUP),PROB)
      KCOM=1
      GO TO 1001
(0    KCOM=0
(1    CALL PLOTHF(IRTPOS+32,LINE+2,5,5HPROB=,PROB,6,3)
      CALL PLOTHI(IRTPOS+ 2,LINE+3,10,10HPREDICTOR ,IPRED(IGROUP),2)
      IPP=IRTPOS+15
      IAL=IPRED(IGROUP)
      DO 1002 L=1,3
      CALL PLOTH(IPP,LINE+3,6,ACR(IAL,L))
1·2   IPP=IPP+6
      CALL PLOTH(IRTPOS+2,LINE+4,5,5HCODES)
      KK=1
      IS2=1
      NC=NCODE(IGROUP)
      IPP=IRTPOS+7
      NVAR=12
      IF(NC.LT.12)NVAR=NC
      NC=NC-NVAR
      LN=LINE+4
      GO TO 1010
033   IF(NC.LE.0)GO TO 1015
      KK=KK+1
      NVAR=14
      IF(NC.LT.14)NVAR=NC
      IPP=IRTPOS+1
      LN=LN+1
      NC=NC-NVAR
C10   DO 1012 L=1,NVAR
      CALL PLOTHI(IPP,LN,0,0,ICODE(IGROUP,IS2),3)
      IPP=IPP+3
C12   IS2=IS2+1
      GO TO (1003,1003,1015),KK
1015  RETURN
      END
```

```
      SUBROUTINE PLEVEL(DF1,DF2,F,P)
      DIMENSION Y(6),ARG(3),GAMA(3)
      BASMLN(Z,X) = (X+.5)*ALOG(X)-X+.918938534+(.833333333E-1-Z*(.27777
     7778E-2-Z*(.793650794E-3-Z*(.595238095E-3-Z*.841750842E-3))))/X
      IF(DF1.LT.1..OR.DF2.LT.1..OR.F.LT.0.)GO TO 14
      IF(ARG(1).EQ.DF1.AND.ARG(2).EQ.DF2)GO TO 7
      ARG(1) = DF1
      ARG(2) = DF2
      ARG(3) = DF1+DF2
      DO 6 I=1,3
      IF(ARG(I).EQ.1.)GO TO 5
      T = ARG(I)-2.
      J = AMOD(ARG(I),2.)+1.
      GO TO (2,1),J
      U = (T-1.)*.5
      GAMA(I) = .572364943-T*.693147181
      IF(U.LT.2.)GO TO 25
      GO TO 3
      T = T*.5
      GAMA(I) = 0.
      IF(T.LT.2.)GO TO 6
      GO TO 4
      Z = 1./(U*U)
      GAMA(I) = GAMA(I)-BASMLN(Z,U)
      Z = 1./(T*T)
      GAMA(I) = GAMA(I)+BASMLN(Z,T)
      GO TO 6
      GAMA(I) = .572364943
      CONTINUE
      C = GAMA(1)+GAMA(2)-GAMA(3)-69.-.693147181
      Y(1) = -.[illegible]4551[illegible]
      Y(2) = -1.203972804
      Y(3) = .5877866649
      Y(4) = Y(2)
      Y(5) = Y(1)
      AX = DF2/(F*DF1+DF2)
      IF (AX.GT.0.99999900) GO TO 136
      H = ATAN(SQRT(AX/(1.-AX)))/60.
      IF(H.LE..130899694E-1)GO TO 8
      H = .261799387E-1-H
      CN = DF1-1.
      CM = (DF2-1.)*.5
      P = -1./H
      XM = -H
      GO TO 9
      CN = DF2-1.
      CM = (DF1-1.)*.5
      P = 0.
      XM = H
      IF (CN.NE.0.) GO TO 95
      XX = Y(2)-C
      IF(XX.LT.0.)GO TO 95
      P=P+EXP(XX-69.)
      Y(6) = -.5108256238
      X = 0.
      DO 13 I=1,10
      GO TO(11,11,11,11,11,11,11,11,11,10),I
      Y(6) = Y(2)
   11 DO 13 J=1,6
      X = X+H
      XS = SIN(X)
      Z = Y(J)-C+CN*ALOG(XS)+CM*ALOG(1.-XS*XS)
      IF(Z)13,12,12
   12 P = P+EXP(Z-69.)
   13 CONTINUE
      P = XM*P
      IF (P.LT.1.0E-05 ) P = 0.
  135 RETURN
  136 P = 1.
      GO TO 135
   14 P = 1.
      PRINT15,DF1,DF2,F
   15 FORMAT(7H1DF1 = F5.0,7H,DF2 = F5.0,5H,F = F10.5//
     148H P DOES NOT EXIST IF DF1 OR DF2 IS LESS THAN 1      /
     248H OR IF F IS LESS THAN ZERO.                         /
     348H P FOR THIS PROBLEM HAS BEEN ABRITRARILY SET        /
     444H EQUAL TO 1. AND A NORMAL RETURN HAS OCCURED.       )
      GO TO 135
      END

      SUBROUTINE PLOTB
      COMMON /PLTARY/ IARY (132,60)
      DATA IBLNK /1H /
      DO 10 I=1,132
      DO 10 J=1,60
   10 IARY(I,J) = IBLNK
      RETURN
      END

      SUBROUTINE PLOTP
      COMMON /PLTARY/ IARY(132,60)
      DATA I1 /1H1/
      IARY(1,1) =I1
      DO 10 J=1,60
   10 PRINT 100, (IARY(I,J), I=1,132)
      RETURN
  100 FORMAT (132A1)
      END

      SUBROUTINE PLOTC (IPRPOS,LINE,ICHAR)
      COMMON /PLTARY/ IARY(132,60)
      IF (IPRPOS .LT. 2) RETURN
      IF (IPRPOS .GT. 132) RETURN
      IF (LINE .LT. 1 ) RETURN
      IF (LINE .GT. 60) RETURN
      IARY(IPRPOS,LINE) = ICHAR
      RETURN
      END
```

```
      SUBROUTINE PLOTH (IPRPOS, LINE,NCHAR,ISTRNG)
      COMMON /WORKA2/ JSTRNG(131)
      CALL STOA1(JSTRNG,ISTRNG)
      CALL PLOTH1(IPRPOS, LINE,NCHAR,JSTRNG)
      RETURN
      END


      SUBROUTINE PLOTV(IPRPOS,LINE,NCHAR,ISTRNG)
      COMMON /WORKA2/ JSTRNG(131)
      CALL STOA1(JSTRNG, ISTRNG)
      CALL PLOTV1(IPRPOS,LINE,NCHAR,JSTRNG)
      RETURN
      END


      SUBROUTINE PLOTH1(IPRPOS,LINE,NCHAR,JSTRNG)
      DIMENSION JSTRNG(131)
      ICHAR = NCHAR
      IF(ICHAR .GT. 131) ICHAR = 131
      DO 10 I=1,ICHAR
      J = I + IPRPOS - 1
   10 CALL PLOTC(J,LINE,JSTRNG(I))
      RETURN
      END


      SUBROUTINE PLOTV1(IPRPOS,LINE,NCHAR,JSTRNG)
      DIMENSION JSTRNG (131)
      ICHAR = NCHAR
      IF (ICHAR .GT. 60) ICHAR = 60
      DO 10 I=1,ICHAR
      J = I + LINE - 1
   10 CALL PLOTC(IPRPOS,J,JSTRNG(I))
      RETURN
      END


      SUBROUTINE STOA1 (JSTRNG, ISTRNG)
      DIMENSION ISTRNG(14), JSTRNG(131)
      DECODE (131,100,ISTRNG) JSTRNG
      RETURN
  100 FORMAT (131A1)
      END


      SUBROUTINE PLOTH1 (IPRPOS,LINE,NCHAR,ISTRNG,IVALUE,IW)
      COMMON /WORKA2/ JSTRNG(131)
      CALL STOA1 (JSTRNG, ISTRNG)
      CALL ITOA1(JSTRNG,NCHAR+1,IVALUE,IW,JW)
      ICHAR = NCHAR + JW
      CALL PLOTH1 (IPRPOS,LINE,ICHAR,JSTRNG)
      RETURN
      END


      SUBROUTINE ITOA1 (ISTRNG,ISC,IVALUE,IW,JW)
      COMMON /WORKA1/ IARY1(10), IARY2(100), IFM(10), LW(2), LD(2)
      DIMENSION ISTRNG(131)
      DATA IBLNK /1H /
      JW = MOD(IW,100)
      IF (JW .GT. 132-ISC) JW = 132-ISC
      ENCODE (2,100,KW) JW
      DECODE (2,102,KW) LW
      I = 3
      IFM(1) = 1H(
      IFM(2) = 1HI
      IF (LW(1) .EQ. IBLNK) GO TO 10
      IFM(3) = LW(1)
      I = 4
   10 IFM(I) = LW(2)
      I = I + 1
      IFM(I) = 1H)
      I = I + 1
      DO 14 J=I,10
   14 IFM(J) = IBLNK
      ENCODE (10,104,IFMT1) IFM
      ENCODE (JW,IFMT1,IARY1) IVALUE
      DECODE (100,106,IARY1) IARY2
      DO 16 I=1,JW
      J = I + ISC - 1
   16 ISTRNG(J) = IARY2(I)
      RETURN
  100 FORMAT (I2)
  102 FORMAT(2A1)
  104 FORMAT(10A1)
  106 FORMAT (100A1)
      END


      SUBROUTINE PLOTHF (IPRPOS,LINE,NCHAR,ISTRNG,FVALUE,IW,ID)
      COMMON /WORKA2/ JSTRNG(131)
      CALL STOA1(JSTRNG,ISTRNG)
      CALL RTOA1(JSTRNG,NCHAR+1,FVALUE,IW,ID,JW)
      ICHAR = NCHAR + JW
      CALL PLOTH1(IPRPOS,LINE,ICHAR,JSTRNG)
      RETURN
      END
```

```
      SUBROUTINE RTOA1 (ISTRNG,ISC,FVALUE,IW,ID,JW)
      COMMON /WORKA1/ IARY1(10), IARY2(108), IFM(6), LW(2), LD(2)
      DIMENSION ISTRNG (131)
      DATA IBLNK /1H /
      JW = MOD(IW,100)
      JC = MOD(ID,100)
      IF (JW .GT. 132-ISC) JW = 132-ISC
      IF (JD .GT. JW-1) JC = JW-1
      ENCODE (2,100,KW) JW
      ENCODE (2,100,KD) JD
      DECODE (2,102,KW) LW
      DECODE (2,102,KD) LD
      I = 3
      IFM(1) = 1H(
      IFM(2) =1HF
      IF (LW(1) .EQ. IBLNK) GO TO 10
      IFM(3) = LW(1)
      I = 4
10    IFM(I) = LW(2)
      I = I + 1
      IFM(I) = 1H.
      I = I + 1
      IF (LD(1) .EQ. IBLNK) GO TO 12
      IFM(I) = LD(1)
      I = I + 1
12    IFM(I) = LD(2)
      I = I + 1
      IFM(I) = 1H)
      I = I + 1
      DO 14 J=I, 10
14    IFM(J) = IBLNK
      ENCODE (10,104,IFM(1)) IFM
      ENCODE (JW,IFM(1),IARY1) FVALUE
      DECODE (100,106,IARY1) IARY2
      DO 16 I=1,JW
      J = I + ISC - 1
16    ISTRNG(J) = IARY2(I)
      RETURN
100   FORMAT(I2)
102   FORMAT (2A1)
104   FORMAT(10A1)
106   FORMAT (100A1)
      END
```

```
      SUBROUTINE AIDPLT(MI,ID,NOGP)
      COMMON/APT1/ D(641,22),ICORH(183),ICORV(183),LEVEL(183)
      DIMENSION MI(183),ID(183),INDEX(183)
      LOGICAL TOP
      INTEGER MI,MI
      DIMENSION K(641,22), HUN(2), TEN(10), UNIT(10)
      EQUIVALENCE (K,D)
      DATA IBL, IASF, AASS, IDOTF, IDOTL, IAS1, AAS2, ARROWT, ARROWB
     1/05555555555555555555,04747474747475555555,04700000000475555555
     2,05757575757575555555,05755555555555555555,00047474747475555555
     3,04711133005547555555,00000000070005555555,00000000071005555555
     4/
      DATA HUN / 00031000000055555555,
     1 00034000000055555555/
      DATA TEN / 00000330000055555555,
     1 00000340000055555555,00000350000055555555,00000360000055555555
     2,00000370000055555555,00000400000055555555,00000410000055555555
     3,00000420000055555555,00000430000055555555,00000440000055555555
     4/
      DATA UNIT / 00000003300055555555,
     1 00000003400055555555,00000003500055555555,00000003600055555555
     2,00000003700055555555,00000004000055555555,00000004100055555555
     3,00000004200055555555,00000004300055555555,00000004400055555555
     4/
      I640 = 640
      DO 10 I = 1, 183
10    LEVEL(I) = 8
      DO 20 I = 1, 641
      DO 20 J = 1, 22
20    K(I, J) = IBL
      ICORH(1) = 321
      ICORV(1) = 2
      K(320, 1) = IAS1
      K(322, 1) = IAS1
      I = 1
      D(321, 1) = AAS2.OR.UNIT(2)
      LEVEL(1) = 1
      DO 70 I = 2, NOGP
      J = ID(I)
      LEVL = LEVEL(J)
      IF (LEVL .GE. 8) GO TO 70
      TOP = MI(J).EQ.I
      ICORHP = ICORH(J)
      ICORVP = ICORV(J)
      ICORV1 = ICORVP+1
      IC = LSHIFT(I640, -(LEVL+1))
      IF (IC .EQ. 2) IC = 1
      IF ( .NOT. TOP) GO TO 40
      IMID = ICORHP-IC
      DO 30 L = IMID, ICORHP
30    K(L, ICORV1) = IDOTL
      GO TO 60
40    IMID = ICORHP+IC
      DO 50 L = ICORHP, IMID
50    K(L, ICORV1) = IDOTL
60    K(ICORHP, ICORVP) = IDOTF
      K(IMID, ICORV1) = IDOTF
      IMIDP1 = IMID+1
```

```
      IMIDM1 = IMID-1
      ICOR1 = ICORV1+1
      K(IMIDP1, ICOR1) = IASF
      K(IMIDM1, ICOR1) = IASF
      II = MOD(I, 10)+1
      JJ = MOD(I/10, 10)+1
      KK = MOD(I/100, 10)+1
      W = AASS.OR.UNIT(II)
      W = W.OR.TEN(JJ)
      W = W.OR.HUN(KK)
      IF (TOP) D(IMID, ICOR1) = W.OR.ARROWT
      IF ( .NOT. TOP) D(IMID, ICOR1) = W.OR.ARROWB
      LEVEL(I) = LEVL+1
      ICORH(I) = IMID
      ICORV(I) = ICORVF+3
 70   CONTINUE
      PRINT 100
      DO 90 I = 1, 641
      TOP = .FALSE.
      DO 80 J = 1, 22
      IF (K(I, J) .NE. IBL .AND. K(I, J) .NE. IDOTL) TOP = .TRUE.
 80   CONTINUE
      IF (TOP) PRINT 110, (K(I, J), J = 1, 22)
 90   CONTINUE
      RETURN
C
100   FORMAT(*1TREE STRUCTURE OF FIRST EIGHT LEVELS, GROUP*,
     1* NUMBER IS GIVEN IN CENTER OF THE BOX, THE PAIR MEMBER*,
     2* WITH THE HIGHER MEAN IS ON TOP*///)
 110  FORMAT  (1X,22A6)
      END
```

CRV

```
      PROGRAM CRV (INPUT,OUTPUT)
      DIMENSION FMT1(22),KFMT1(22),FMT2(22),KFMT2(22)
      DIMENSION A(35000),KA(35000)
      COMMON FMT1,FMT2,A
      EQUIVALENCE (FMT1,KFMT1),(FMT2,KFMT2),(A,KA)
C STORAGE = NR(3NV+2) + NV(NT+NV+10) + 2NT + 500
      READ 5, NR, NC, NV, NT, NMOD, PN, PD
    5 FORMAT ( 5I5, 2F5)
         P = PN / PD
      NV2= NV*2 $ K=NV2
      KFMT2(22) =201 $ CALL DEF(LF,NR,NV) $ CALL DEF(LF1,NR,NV)
      CALL DEF(LF2,NR,NV)$CALL DEF (LT,NT,NV2)$LM=KFMT2(22)$LS=LM+K
      LR=LS+K $ LB=LR+K*K$LW=LB+NV$LRSQ=LW+NV+1$LY=LRSQ+NV+248$LP=LY+NR
      CALLFRMT (KFMT1)
      CALL CDTOCR(LF,NR,NC,NV,1,101)
      KFMT2(20) = 0
        KFMT2(17) = 3
      KFMT2(17)=1
      DO 68 MOD = 1, NMOD
   67 DO 10 N = 1, NT
      CALL SCATR (LF,NR,NV,LF1,NR1,LF2,NR2,P)
      PRINT 22, NR1, NR2
   22 FORMAT (5X, 2I10)
      CALL CORRLJ (NR1,NV,LM,LS,-LR,1,101,NV, 0,LF1)
       CALL REGREJ (LM,LS,LR,LB,LW,LRSQ,NV,NR1,5,.000001)
      A(LRSQ+247) = 0.
      FMT2(11)=0.
      CALL PREDICT(6,NR2,NV,NV,LW,LRSQ+200,1,101,LF2,LP,LY,0,2)
      CALL RSQ (LY,LP,NR2,R )
      CALL STORE (LW+1,NV+1,1,LT,NV+2, N)
      L =  (N-1)*(NV+2) + LT + NV+1
   10 A(L) = R
      CALL PRINT (LT,NT,NV+2, 1, 0)
      CALL CORRLJ(NT,NV2,LM,LS,-LR,1,101,NV2,6,LT)
      CALL PREDICT(6,NR ,N     ,LM,LRSQ+200,1,101,LF ,LP,LY,0,1)
      CALL TABDEV(LP,LY,NC   ,F,CUT)
       CALL HITS(LY,LP,NR ,CUT)
   68 CONTINUE
   17 END

      SUBROUTINE DATMAN (I,NC,J,NV) $ COMMON B(46),C(100),V(100)
      CALL ZEROST (J+1, 1, NV)
        K=C(1)
       V(K)=1
      V(13)=C(5)
   99 RETURN $ END
```

NCRIT

```
JOB,TM=25,PR=25.
RUN(S)
CP.
READPF(2350,EOSTATJ)
RFL,77777.
NOREDUCE.
LOAD(LGO,EOSTATJ)
EXECUTE.
      PROGRAM DRIVER(INPUT,OUTPUT,TAPE3)
      DIMENSION FMT1(22),KFMT1(22),FMT2(22),KFMT2(22),A(10000),KA(10000)
      COMMON FMT1,FMT2,A
      EQUIVALENCE(FMT1,KFMT1),(FMT2,KFMT2),(A,KA)
      CALL TAPGEN(3)
      CALL ALPHA(KFMT1,6,6H(8A10))
      CALL CORXTP(3,224,8,0,1,9,63,18,13,13HDATA FILE ONE)
      CALL RFGWTS(17,16,601)
      CALL TABLES(3,10HDATA FILE ,224,17,17,601,650,2)
      CALL EXIT
      END
            SUBROUTINE DATRAM (I,NC,J,NV) $ DIMENSION C(50),V(50)
      DIMENSION FMT1(22),KFMT1(22),FMT2(22),KFMT2(22),A(1),KA(1)
      COMMON FMT1,FMT2,A
      EQUIVALENCE (FMT1,KFMT1),(FMT2,KFMT2),(A,KA)
      CALL ZFROST (J+1, 1, NV)
      DO 6 M=1,50
6     V(LM)=0.
      DO 8 L=1,NC
 8    C(L)=A(I+L)
      DO 9 N=1,NC
      V(N)= C(N+1)
9     V(17)=C(1)
      DO 10 L=1,NV
      L1=L
10    A(J+L)=V(L1)
   99 RETURN $ END
      NOLIST
      SUBROUTINE YMT(N,TO,FROM)
      DIMENSION TO(N),FROM(N)
      DO 5 I=1,N
5     TO(I)=FROM(I)
      RETURN  $END
      SUBROUTINE REGWTS(NVARS,NPREDS,LRGWTS)
C     GENERATES REGRESSION WEIGHTS PUNCHED FROM THILON INTO AN ARRAY
C     STARTING IN LOCATION A(LRGWTS)
C     ARGUMENTS
C     NVARS -NUMBER OF VARIABLES IN CORRELATION MATRIX FROM REGRESSION PHASE
C     NPREDS-NUMBER OF PREDICTOR VARIABLES IN REGRESSION PHASE
C             (NORMALLY NPREDS IS ONE LESS THAN NVARS UNLESS SOME PREDICTOR
C             VARIABLES ARE NOT USED IN REGRESSION PHASE)
C     LRGWTS-A VECTOR OF NVARS+1 RAW SCORE REGRESSION WEIGHTS.ELEMENTS OF THIS
C             VECTOR CORRESPONDING TO VARIABLES NOT SPECIFIED AS PREDICTORS IN
C             A PARTICULAR PROBLEM AND TO THE CRITERION VARIABLE WILL BE ZERO.
C             THE LAST ELEMENT IS THE REGRESSION CONSTANT
C     ORDER OF INPUT CARDS
C           1-NPREDS RAW SCORE REGRESSION WEIGHT CARDS
C             TRICOR FORMAT(16X,I4,F15.8),PREDICTOR NUMBER,RAW WEIGHT
C           2-ONE REGRESSION CONSTANT CARD
C             READ WITH FORMAT(20X,F15.8)
C           3-ONE F CARD
C             TRICOR FORMAT(2A6,I3,I5,I3,F12.8)STUDY NUMBER,PHASE NUMBER OF
C             REGRESSION PHASE,PROBLEM NUMBER,SAMPLE N,NUMBER OF PREDICTORS,RSQ
```

```
      DIMENSION FMT1(22),KFMT1(22),FMT2(22),KFMT2(22),A(1),KA(1)
      COMMON FMT1,FMT2,A
      EQUIVALENCE (FMT1,KFMT1),(FMT2,KFMT2),(A,KA)
      COMMON /ABC/STNUM,PHNUM,NUMPB,NUMSAM,NUMPRD,RSQ,NRV1
      A(LRGWTS)=0.0
      CALL TMT(NVARS,A(LRGWTS+1),A(LRGWTS))
      DO 10 I=1,NPREDS
      READ 500,IDPRED,WT
  500 FORMAT(16X,I4,F15.8)
C     ISUB=LRGWTS+IDPRED-1
C     ISUB=LRGWTS+I-1
      ISUB=LRGWTS+IDPRED-1
   10 A(ISUB)=WT
      READ 502,REGCON
  502 FORMAT(20X,F15.8)
      LSUB=LRGWTS+NVARS
      A(LSUB)=REGCON
C     READ F CARD
      READ 501,STNUM,PHNUM,NUMPB,NUMSAM,NUMPRD,RSQ,NRV1
C     NUMPRD=NUMPRD+1
  501 FORMAT(2A6,I3,I5,I3,F12.8,I3)
      READ 6,(KFMT2(JLK),JLK=1,8)
6     FORMAT(8A10)
      RETURN
      END
$IBFTC TABLES
      SUBROUTINE TABLES(IDTAPE,FILE,NCASES,NVARS,NCRIT,LRGWTS,LWORK,N)
C     GENERATES PREDICTION SYSTEM TABLES 1 AND 2
C     ARGUMENTS
C     IDTAPE-TAPE DEVICE TO BE READ(0-4)
C     FILE  -NAME OF PERSUB BINARY TYPE FILE TO BE READ(12H...)
C     NCASES-NUMBER OF CASES TO BE READ
C     NVARS -NUMBER OF VARIABLES TO BE READ,INCLUDING CRITERION VARIABLE
C     NCRIT -VARIABLE NUMBER OF THE CRITERION VARIABLE(ACTUAL VALUE)
C     LRGWTS-A VECTOR OF NVARS+1 RAW SCORE REGRESSION WEIGHTS.ELEMENTS OF THIS
C            VECTOR CORRESPONDING TO VARIABLES NOT SPECIFIED AS PREDICTORS IN
C            A PARTICULAR PROBLEM AND TO THE CRITERION VARIABLE WILL BE ZERO.
C            THE LAST ELEMENT IS THE REGRESSION CONSTANT
C     LWORK -A VECTOR OF WORK AREA AT LEAST NVARS+1111 ELEMENTS
C     N      -THE PRINT OPTION
C              1-PRINT PERCENT TABLE ONLY
C              2-PRINT BOTH PERCENT AND FREQUENCY TABLE
      DIMENSION FMT1(22),KFMT1(22),FMT2(22),KFMT2(22),A(1),KA(1)
      COMMON FMT1,FMT2,A
      EQUIVALENCE (FMT1,KFMT1),(FMT2,KFMT2),(A,KA)
      DIMENSION FILE(2),KPRINT(11)
      COMMON /ABC/STNUM,PHNUM,NUMPB,NUMSAM,NUMPRD,RSQ,NRV1
      CALL POSTAP(LOCREC,IDTAPE,0,FILE)
      LTPIN=LWORK
      LTPEND=LWORK+NVARS-1
      LONES=LTPEND+1
      LZEROS=LONES+101
      LINT=LZEROS+101
      LP1A1=LINT+101
      LP0A0=LP1A1+101
      LP1A0=LP0A0+101
      LP0A1=LP1A0+101
      LP1=LP0A1+101
      LP0=LP1+101
      LNCORR=LP0+101
      LOF=LNCORR+101
      LEND=LOF
C     CLEAR STORAGE
      CALL ZEROST(LWORK+1,LEND)
      SUMX=0.0
      SUMY=0.0
      SUMX2=0.0
      SUMY2=0.0
      SUMXY=0.0
      TOTAL0=0.0
      TOTAL1=0.0
      NRV=NRV1+3000
      NOREC=0
      DO 805 ILOOP=1,NCASES
C     READ(IDTAPE)KSKIP,(A(IT),IT=LTPIN,LTPEND)
      CALL RDBKTP(NOREC,3,63,8,2000,10,10HDATA FILE )
      DECODE(80,KFMT2,A(2000))     (A(JI),JI=3000,NRV)
      CALL DATRAN(2999,NRV1,LTPIN-1,NVARS)
      SUMLC=0.0
      DO 801 ILC=1,NVARS
      I1=LRGWTS+ILC-1
      I2=LTPIN+ILC-1
  801 SUMLC=SUMLC+A(I1)*A(I2)
      I3=LRGWTS+NVARS
      SUMLC=SUMLC+A(I3)
      IF(ILOOP.GT.1)GO TO 550
      PRINT 540
540   FORMAT(*1  VARIABLE      RAW WEIGHT*)
      DO 551 IW=1,NVARS
      I1=LRGWTS+IW-1
      PRINT 552,IW,A(I1)
552   FORMAT(1X,I10,F15.8)
551   CONTINUE
      PRINT 553,A(I3)
  553 FORMAT(//*    REG CON*,F15.8)
      PRINT 560,NCRIT
  560 FORMAT(//*   CRIT VAR*,I6/*1RAW DATA*)
      PRINT 6, KFMT2
6     FORMAT(8A10)
550   IF(ILOOP.GE. 2)GO TO 554
      PRINT 555,(A(IW),IW=LTPIN,LTPEND),SUMLC
555   FORMAT(/1X,12F10.3)
554   CONTINUE
      I4=LTPIN+NCRIT-1
      ACTUAL=A(I4)
C   SPECIAL LOGIC FOR DATA CODED AS 1  AND 2  FOR BINARY 0 AND 1
         ACTUAL=A(I4)-1
      IF(NUMPR.GE.10) GO TO 1702
      GO TO 1701
C SPECIAL PRINT OF RESIDUAL TABLE
1702    REST=ACTUAL-SUMLC
      IF(ILOOP.EQ.1)PRINT 965
965   FORMAT(35X*RESIDUAL TABLE*/10X*CASE*10X*ACTUAL*7X*PREDICTED*6X*RES
     1IDUAL*)
      PRINT 964,ILOOP,ACTUAL,SUMLC,REST
964   FORMAT(17X,I5,5X,3(4X,F7.4,4X))
1701    CONTINUE
      SUMX=SUMX+SUMLC
      SUMX2=SUMX2+SUMLC**2
      SUMY=SUMY+ACTUAL
      SUMY2=SUMY2+ACTUAL**2
      SUMXY=SUMXY+SUMLC*ACTUAL
C     DETERMINE INTERVAL FOR SUMLC
      IF(SUMLC.GE.1.00)GO TO 901
```

```
      IF(SUMLC.LT..01)GO TO 902
      SUM=SUMLC*100.0
      ISUB=SUM
      I4=LINT+ISUB
      A(I4)=A(I4)+1.0
      GO TO 903
  901 ISUB=100
      A(LINT+100)=A(LINT+100)+1.0
      GO TO 903
  902 ISUB=0
      A(LINT)=A(LINT)+1.0
  903 IF(ACTUAL.EQ.0.0)GO TO 904
C     ACTUAL SCORE=1.0
      TOTAL1=TOTAL1+1.0
      I5=LP1A1+ISUB
      A(I5)=A(I5)+1.0
      I6=LONES+ISUB
      A(I6)=A(I6)+1.0
      GO TO 800
C     ACTUAL SCORE=0.0
  904 TOTAL0=TOTAL0+1.0
      I7=LP0A0+ISUB
      A(I7)=A(I7)+1.0
      I8=LZEROS+ISUB
      A(I8)=A(I8)+1.0
  800 CONTINUE
      REWIND IOTAPE
C     ACCUMULATE P1A1 DOWN AND P0A0 UP
      DO 820 K=1,100
      K1=LP1A1+100-K
      A(K1)=A(K1)+A(K1+1)
      K5=LP0A0+K
      A(K5)=A(K5)+A(K5-1)
  820 CONTINUE
      DO 825 KE=1,100
      KZ=LP0A0+101-KE
  825 A(KZ)=A(KZ-1)
      A(LP0A0)=0.0
      GENERATE P1A0 AND P0A1
      DO 802 IA=1,101
      I9=LP1A0+IA-1
      I10=LP0A0+IA-1
      A(I9)=TOTAL0-A(I10)
      I11=LP0A1+IA-1
      I12=LP1A1+IA-1
      A(I11)=TOTAL1-A(I12)
  802 CONTINUE
C     GENERATE LP1 AND LP0
      DO 804 IC=1,101
      K1=LP1A1+IC-1
      K2=LP1A0+IC-1
      K3=LP1+IC-1
      A(K3)=A(K1)+A(K2)
      K4=LP0A0+IC-1
      K5=LP0A1+IC-1
      K6=LP0+IC-1
      A(K6)=A(K4)+A(K5)
  804 CONTINUE
C     GENERATE LNCORR AND LOF
      DO 821 IC=1,101
      K1=LP1A1+IC-1
      K2=LP1A0+IC-1
      K4=LP0A0+IC-1
      K5=LP0A1+IC-1
      K7=LNCORR+IC-1
      K8=LOF+IC-1
      A(K7)=A(K1)+A(K4)
      CALL OBJFUN(A(K1),A(K2),A(K4),A(K5),A(K8))
  821 CONTINUE
C     PRINT TITLES
      PRINT 604
  604 FORMAT(34H1THE PREDICTION OF A BINARY VECTOR,31X,64H
     1                  PERSONNEL DIVISION,LACKLAND AFB,TEXAS )
      PRINT 606,NUMPB,STNUM,PHNUM,NUMSAK,NUMPRD
  606 FORMAT(19H PREDICTION SYSTEM=,I2,11H SAMPLE IU=,2A6,13H N OF SAMPL
     1E=,I4,17H N OF PREDICTORS=,I4,47H OBJECTIVE FUNCTION=V11*C1+V10*C2
     2+V00*C3+V01*C4 )
      PRINT 605,FILE(1),FILE(2),NCASES
  605 FORMAT(19H APPLIED SAMPLE IU=,2A6,13H N OF SAMPLE=,I5/)
      PRINT 607
  607 FORMAT(47X,35HTABLE TYPE 1-CUMULATIVE PERCENTAGES/)
      PRINT 700
  700 FORMAT(12H PREDICTED  ,12H PERCENT OF ,12H PERCENT OF ,12H PERCENT
     1 OF ,12H PERCENT OF ,12H PERCENT OF ,12H PERCENT OF ,12H PERCENT O
     2F ,12H PERCENT OF ,12H PERCENT OF ,12H  OBJECTIVE )
      PRINT 701
  701 FORMAT(12H  SCORE     ,12H TOTAL   IN ,12H NUM IN INT ,12H   PRED
     11   ,12H   PRED 1   ,12H   TOTAL    ,12H   PRED 0   ,12H   PRED 0
     2 ,12H   TOTAL    ,12H   TOTAL    ,12H  FUNCTION  )
      PRINT 702
  702 FORMAT(12H  CUT-OFF   ,12H  INTERVAL  ,12H IN   ACT 1 ,12H   ACT
     11   ,12H   ACT 0    ,12H   PRED 1   ,12H   ACT 1    ,12H   ACT 0
     2 ,12H   PRED 0   ,12H  CORRECT  /)
C     PRINT TABLE 1
      PINT=1.01
      NLINE=0
      GTOTAL=TOTAL1+TOTAL0
      DO 803 IB=1,101
      J1=LONES+101-IB
      J2=LZEROS+101-IB
      J3=LINT+101-IB
      J4=LP1A1+101-IB
      J5=LP0A0+101-IB
      J6=LP1A0+101-IB
      J7=LP0A1+101-IB
      J8=LP1+101-IB
      J9=LP0+101-IB
      J10=LNCORR+101-IB
      J11=LOF+101-IB
      PINT=PINT-.01
      NLINE=NLINE+1
      IF(A(J3).EQ.0) GO TO 974
      IF(A(J8).EQ.0) GO TO 974
      IF(A(J9).EQ.0) GO TO 974
      KPRINT(2)=((A(J3)/GTOTAL)+.005)*100.0
      KPRINT(3)=((A(J1)/A(J3))+.005)*100.0
      KPRINT(4)=((A(J4)/A(J8))+.005)*100.0
      KPRINT(5)=((A(J6)/A(J8))+.005)*100.0
      KPRINT(6)=((A(J8)/GTOTAL)+.005)*100.0
      KPRINT(7)=((A(J7)/A(J9))+.005)*100.0
      KPRINT(8)=((A(J5)/A(J9))+.005)*100.0
      KPRINT(9)=((A(J9)/GTOTAL)+.005)*100.0
      KPRINT(10)=((A(J10)/GTOTAL)+.005)*100.0
      KPRINT(11)=A(J11)
```

```
      PRINT 806,PINT,(KPRINT(L),L=2,11)
  806 FORMAT(1X,F6.2,5X,10(I8,4X))
      GO TO 973
 974  PRINT 975,PINT
 975  FORMAT(6 THERE ARE NULL CATEGORIES FOR THIS INTERVAL *F6.2)
 973  IF(NLINE.NE.48)GO TO 803
      PRINT 807
  807 FORMAT(12H1PREDICTED  ,12H PERCENT OF ,12H PERCENT OF ,12H PERCENT
     1 OF ,12H PERCENT OF ,12H PERCENT OF ,12H PERCENT OF ,12H PERCENT O
     2F ,12H PERCENT OF ,12H PERCENT OF ,12H  OBJECTIVE )
      PRINT 808
  808 FORMAT(12H  SCORE     ,12H TOTAL   IN ,12H NUM IN INT ,12H   PRED
     11   ,12H  PRED 1   ,12H   TOTAL    ,12H   PRED 0   ,12H   PRED 0
     2 ,12H   TOTAL    ,12H   TOTAL    ,12H  FUNCTION  )
      PRINT 809
  809 FORMAT(12H  CUT-OFF   ,12H  INTERVAL  ,12H IN   ACT 1 ,12H   ACT
     11   ,12H   ACT  0   ,12H   PRED 1   ,12H   ACT  1   ,12H   ACT  0
     2 ,12H   PRED 0   ,12H  CORRECT   /)
  803 CONTINUE
C COMPUTE THE MEAN,SD,RSQ
C LET PRED SCORES=X,ACTUAL SCORES=Y
      CASES=NCASES
      XMEAN=SUMX/CASES
      YMEAN=SUMY/CASES
      XSD=(SQRT(CASES*SUMX2-(SUMX**2)))/CASES
      YSD=(SQRT(CASES*SUMY2-(SUMY**2)))/CASES
      RSQXY=((CASES*SUMXY-SUMX*SUMY)**2)/((CASES*SUMX2-SUMX**2)*(CASES
     1*SUMY2-SUMY**2))
      PRINT 506,XMEAN,YMEAN,RSQXY
  506 FORMAT(/31H CHECK ANALYSIS-PREDICTED MEAN=,F12.4,5X,
     112HACTUAL MEAN=,F12.4,5X,6HRSQPA=,F12.8)
      IF(N.EQ.1)RETURN
C     PRINT TITLES
      PRINT 706
  706 FORMAT(1H1,44X,36H TABLE 2-CUMULATIVE FREQUENCY COUNTS/)
      PRINT 703
  703 FORMAT(12H PREDICTED  ,12H   NONCUM   ,12H   NONCUM   ,12H   NUMBE
     1R   ,12H   NUMBER   ,12H   NUMBER   ,12H   NUMBER   ,12H   NUMBER
     2 ,12H   NUMBER   ,12H   NUMBER   ,12H  OBJECTIVE )
      PRINT 704
  704 FORMAT(12H   SCORE    ,12H   NUM IN   ,12H ACT 1 IN  ,12H   PRED
     11   ,12H   PRED 1   ,12H   PRED 1   ,12H   PRED 0   ,12H   PRED 0
     2 ,12H   PRED 0   ,12H   CORRECT  ,12H  FUNCTION  )
      PRINT 705
  705 FORMAT(12H  CUT-OFF   ,12H  INTERVAL  ,12H  INTERVAL  ,12H   ACT
     11   ,12H   ACT  0   ,12X,12H   ACT  1   ,12H   ACT  0   /)
C     PRINT TABLE 2
      PINT=1.01
      NLINE=0
      DO 710 IB=1,101
      J1=LONFS+101-IB
      J2=LZEROS+101-IB
      J3=LINT+101-IB
      J4=LP1A1+101-IB
      J5=LP0A0+101-IB
      J6=LP1A0+101-IB
      J7=LP0A1+101-IB
      J8=LP1+101-IB
      J9=LP0+101-IB
      J10=LNCORR+101-IB
      J11=LOF+101-IB
      PINT=PINT-.01
      NLINE=NLINE+1
      KPRINT(2)=A(J3)
      KPRINT(3)=A(J1)
      KPRINT(4)=A(J4)
      KPRINT(5)=A(J6)
      KPRINT(6)=A(J8)
      KPRINT(7)=A(J7)
      KPRINT(8)=A(J5)
      KPRINT(9)=A(J9)
      KPRINT(10)=A(J10)
      KPRINT(11)=A(J11)
      PRINT 806,PINT,(KPRINT(L),L=2,11)
      IF(NLINE.NE.51)GO TO 710
      PRINT 740
  740 FORMAT(12H1PREDICTED  ,12H   NONCUM   ,12H   NONCUM   ,12H   NUMBE
     1R   ,12H   NUMBER   ,12H   NUMBER   ,12H   NUMBER   ,12H   NUMBER
     2 ,12H   NUMBER   ,12H   NUMBER   ,12H  OBJECTIVE )
      PRINT 741
  741 FORMAT(12H   SCORE    ,12H   NUM IN   ,12H ACT 1 IN  ,12H   PRED
     11   ,12H   PRED 1   ,12H   PRED 1   ,12H   PRED 0   ,12H   PRED 0
     2 ,12H   PRED 0   ,12H   CORRECT  ,12H  FUNCTION  )
      PRINT 742
  742 FORMAT(12H  CUT-OFF   ,12H  INTERVAL  ,12H  INTERVAL  ,12H   ACT
     11   ,12H   ACT  0   ,12X,12H   ACT  1   ,12H   ACT  0   /)
  710 CONTINUE
      RETURN
      END
$IBFTC OBJFUN
      SUBROUTINE OBJFUN(ARG1,ARG2,ARG3,ARG4,ALOC)
C ARG1=PRED1-ACT1
C ARG2=PRED1-ACT0
C ARG3=PRED0-ACT0
C ARG4=PRED0-ACT1
      DIMENSION FMT1(22),KFMT1(22),FMT2(22),KFMT2(22),A(1),KA(1)
      COMMON FMT1,FMT2,A
      EQUIVALENCE(FMT1,KFMT1),(FMT2,KFMT2),(A,KA)
C DEFINE CONSTANTS
      CON1=1.0
      CON2=0.0
      CON3=1.0
      CON4=0.0
      ALOC=ARG1*CON1+ARG2*CON2+ARG3*CON3+ARG4*CON4
      RETURN
      END
```

References

Anderberg, M. R., Cluster Analysis for Applications, Ph. D. Dissertation, The University of Texas at Austin, December, 1971.

Jennings, E., EDSTAT-J: A Subroutine System for Data Processing in the Behavioral Sciences, Department of Educational Psychology, University of Texas at Austin, 1971.

Koylyay, J. B., C. D. Gott, and J. H. Elton, Automatic Interaction Detection AID-4, Personnel Research Div., AFHRL, Lackland AFB, Texas, July, 1971, (Personal Communication).

SOCLIB Program Manual (Program MCA), The University of Texas at Austin Computational Center.

Step-01; Statistical Computer Program for Stepwise Multiple Regression, Center for Highway Research, The University of Texas at Austin, 1970.

BIBLIOGRAPHY

1. Allais, D. C., The Selection of Measurements for Predictions, Tech. Report No.-6103-9, Stanford Electronics Laboratories, Stanford, California, November 1964.

2. Anderberg, M. R., Cluster Analysis for Applications, Unpublished Ph.D. Dissertation, The University of Texas at Austin, December 1971.

3. Anderson, R. L. and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill, 1952.

4. Andrews, F., Morgan, J., and J. Sonquist, Multiple Classification Analysis, ISR, University of Michigan, Ann Arbor, 1967.

5. Beach, Lee Roy, "Multiple Regression as a Model for Human Information Utilization", Organizational Behavior and Human Performance, Vol. 2, p. 276-289, 1967.

6. Bettman, James R., Applying a New Methodological Approach to a Problem of Information Processing Model Validation, Working Paper No. 152, Western Management Science Institute, UCLA, 1969.

7. Bettman, James R., Behavior Simulation Models in Marketing Systems, Unpublished Dissertation, Yale University, Boston, Massachusetts, 1969.

8. Biborosch, R. A., "Numerical Credit Scoring", Credit World, Vol. 53: no. 9 & 10, June-July 1965.

9. Blaylock, H. M. Jr., Theory Construction, Prentice-Hall, 1969.

10. Bottenberg, R. A. and R. E. Christal, An Iterative Technique for Clustering Criteria which Retains Optimum Predictive Efficiency, WADD-TN-61-30, Personnel Research Laboratory, Lackland, Air Force Base, Texas, March 1961.

11. Bottenberg, R. A. and Joe H. Ward, Jr., Applied Multiple Linear Regression, PRL-TDR-63-6, 6470th Personnel Research Laboratory, Lackland Air Force Base, Texas, March 1963.

12. Bowman, E. H., "Consistency and Optimality in Managerial Decision Making", Management Science, No. 9, p. 310-321, 1963.

13. Brehmer, B., "The Structure of Policy Conflicts", UMEA Psychological Reports, No. 29, Department of Psychology, University of UMEA, Sweden, 1970.

14. Bruner, J. S.; Goodnow, J. J. and G. A. Austin, A Study in Thinking, Wiley and Sons, New York, 1956.

15. Brunswik, Egon, "Thing Constancy as Measured by Correlation Coefficients", Psychological Review, Vol. 47, p. 69-78, 1940.

16. Brunswik, E., Perception and the Representative Design of Experiments, University of California Press, Berkeley, 1956.

17. Christai, R. E., "JAN: A Technique for Analyzing Group Judgment", The Journal of Experimental Education, Vol. 36, NR 4, p. 24-27, AD 675295, Summer 1968 (issued earlier under the same title as report No. PRL-TDR-63-3, 6570th Personnel Research Laboratory, Lackland AFB, Texas, AD 403813, February 1963).

18. Christal, R. E., "Selecting a Harem and Other Applications of the Policy Capturing Model", PRL-TR-67-1, AFHRL, Lackland AFB, Texas, March 1, 1967.

19. Churchman, C. W., Prediction and Optimal Decisions: Philosophical Issues in a Science of Values, Prentice-Hall, 1961.

20. Cohen, Kalman J., and Frederick S. Hammers, Eds., Analytical Methods in Banking, Irwin, 1966.

21. Credit Research Foundation, Credit Limits Established by Formula and Computer, September 1970.

22. Dalkey, Norman C., and Daniel L. Rourke, Experimental Assessment of Delphi Procedures with Group Value Judgments, R-612-ARPA, RAND Corp., Santa Monica, California, February, 1971.

23. Dawes, Robyn M., and Richard D. Diller, "The Prediction of Bootstrapping and a Method of Amalgamation", Oregon Research Institute Research Bulletin, Vol. 10, No. 6, October 1970.

24. Dawes, Robyn M., "A Case Study of Graduate Admission: Application of Three Principles of Human Decision Making", American Psychologist, Vol. 26, No. 2, February 1971.

25. Draper, N. R., and H. Smith, Applied Regression Analysis, Wiley, 1966.

26. Dudycha, D. S., and Naylor, J. C., The Effect of Variations in the Cue R Matrix Upon the Obtained Policy Equation of Judges", Educational and Psychological Measurement, Vol. 26, p. 583-603.

27. Durand, David, Risk Elements in Consumer Installment Financing, Study #8, National Bureau of Economic Research, New York, 1941.

28. Earle, T. C., "Task Learning, Interpersonal Learning and Cognitive Complexity", Program on Cognitive Processes Report, No. 129, Institute of Behavorial Science, University of Colorado, Boulder, May 1970.

29. Einhorn, Hillel J., "The Use of Non-Linear, Non-Compensatory Models in Decision Making", Psychological Bulletin, Vol. 3, No. 3, p. 221-230, 1970.

30. Einhorn, Hillel J., "Use of Non-Linear, Non-Compensatory Models as a Function of Task and Amount of Information", Organizational Behavior and Human Performance, Vol. 6, p. 1-27, January 1971.

31. Einhorn, Hillel J., Expert Measurement and Mechanical Combination, II: An Analysis of the Validity, Reliability and Agreement of Expert Judges, Working Paper 21-71-2, Carnegie Mellon University, September 13, 1971.

32. Einhorn, Hillel J., "Expert Measurement and Mechanical Combination", Organizational Behavior and Human Performance, March 1972.

33. FARADA Handbooks, Vol. 1 through 6, Naval Ordinance System Command, FMSAEG, Corona, California, April 1969.

34. Finifter, B. M., "ERIV: A Computer Program for Evaluating Relative Importance of Variables in the Analysis of Interaction Effects", Behavioral Science, Vol. 16, No. 5, pp. 511-512, September 1971.

35. Finifter, Bernard M., Program ERIV and PEP Statistic for Measuring Relative Importance, personal correspondence, May 1972.

36. Fishburn, P. C., "Utility Theory", Management Science, No. 14, p. 335-378, 1968.

37. Fishburn, P. C., Utility Theory for Decision Making, Wiley & Sons, New York, 1970.

38. Garner, W. R., "The Stimulus in Information Processing", American Psychologists, Vol. 25, p. 350-358, 1970.

39. Glasser, Gerald J., "Statistical and Mathematical Application in Consumer Credit Management", Credit World, Vol. 55, No. 1 and 2, p. 18-21, October - November 1966.

40. Goldberg, L. R., "Simple Models or Simple Processes? Some Research on Clinical Judgments", American Psychologist, Vol. 23, p. 483-496, 1968.

41. Goldberg, L. R., "Man Versus Model of Man: A Rationale plus Some Evidence for a Method of Improving on Clinical Inferences", Psychological Bulletin, Vol. 73, p. 422-432, 1970.

42. Goldberg, Lewis R., "Five Models of Clinical Judgment: An Empirical Comparison between Linear and Non-Linear Representation of the Human Inference Process", Organizational Behavior and Human Performance, Vol. 6, p. 458-479, 1971.

43. Goodman, Leo A., "The Multivariate Analysis of Qualitative Data: interactions Among Multiple Classifications", Journal of American Statistical Association, Vol. 65, p. 226-256, March 1970.

44. Graybill, J. A., An Introduction to Linear Statistical Models, McGraw-Hill, New York, 1961.

45. Green, Bert F., "Description and Explanations: A Comment on Papers by Hoffman and Edwards", in Formal Representation of Human Judgment, B. Kleinmuntz (ed.), Wiley, 1968.

46. Grinnell, Merilyn; Keeley, Stuart; and Michael Doherty, "Bayesian Predictions of Faculty Judgments of Graduate School Success", Organizational Behavior and Human Performance, Vol. 6, p. 379-387, May 1971.

47. Hammond, K. R., Hursch, C. J., and F. J. Todd, "Analyzing the Components of Clinical Inference", Psychological Review, Vol. 71, p. 438-456, 1964.

48. Hayes, John R., "Strategies in Judgmental Research", in Formal Representation of Human Judgment, B. Kleinmuntz (ed.), Wiley, 1968.

49. Helmer, Olaf, Analysis of the Future: The Delphi Method, RAND Corp., Santa Monica, Calif., March 1967.

50. Herzberg, Paul A., The Parameters of Cross Validation, Ph.D. Dissertation, University of Illinois, NTIS AD #667467, Dec., 1967.

51. Hicks, Charles, R., Fundamental Concepts in the Design of Experiments, Holt, Rinehart, and Winston, New York, 1964.

52. Hoffman, Paul J., "The Paramorphic Representation of Clinical Judgment", Psychological Bulletin, Vol. 47, p. 116-131, 1960.

53. Hoffman, Paul J., "Assessment of the Independent Contributions of Predictors", Psychological Bulletin, Vol. 59, No. 1, pp. 77-80, 1962.

54. Hoffman, Paul J., "Cue Consistency in Human Judgment", in Formal Representation of Human Judgment, B. Kleinmuntz (ed.), Wiley, 1968.

55. Hoffman, Paul J., P. Slovic, and L. G. Rorer, "An Analysis-of-Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment", Psychological Bulletin, Vol. 69, p. 338-349, 1968.

56. Holt, Robert R., "Yet Another Look at Clinical and Statistical Prediction: Or, Is Clinical Psychology Worthwhile?", American Psychologist, Vol. 25, p. 337-349, 1970.

57. Huber, G. P.; Daneshgor, R., and D. L. Ford, "An Empirical Comparison of Five Utility Models for Predicting Job Preferences", Organizational Behavior and Human Performance, Vol. 6, p. 267-282, May 1971.

58. Jennings, E., "Fixed Effects Analysis of Variance by Regression Analysis", Multivariate Behavorial Research, Vol. 2, p. 95-108, 1967.

59. Jennings, E., EDSTAT-J: A Subroutine System for Data Processing in the Behavorial Sciences, Department of Educational Psychology, University of Texas at Austin, Feb., 1971.

60. Kempthorne, O., The Design and Analysis of Experiments, New York, Wiley & Co., 1952.

61. Kennedy, Henry A., Judged Importance of the Dimensions of A Loan, University of Washington, Department of Psychology, May 1971.

62. Klahr, David, "Decision Making in a Complex Environment: The Use of Similarity Judgments to Predict Preferences", Management Science, Vol. 15, p. 595-609, July 1969.

63. Kleinmuntz, B. (ed.), Formal Representation of Human Judgment, Carnegie-Mellon University, Wiley, 1968.

64. Koplyay, J. B., Field Test of the Weighted Airman Promotion System: Phase I. Analysis of the Promotion Board Component in the Weighted Factors System, Report No. AFHRL-TR-69-101, AF Human Resources Laboratory, 20 pp., AD 689751, April 1969.

65. Koplyay, J. B., Field Test of the Weighted Airman Promotion System: Phase II. Validation of the System for Grades E-4 Through E-7., Report No. AFHRL-TR-69-102, 18 pp., AD 697798, May 1969.

66. Koplyay, James B., etal., "Automatic Interaction Detection AID-4, Personnel Division, AFHRL, Lackland Air Force Base, Texas, (personal communication, 1971).

67. Krieger, Sam, "Appraising Credit Applications by Point Count Method", Credit World, Vol. 53, No. 6, p. 18-20, March 1965.

68. Kunreuther, Howard, "Extensions of Bowman's Theory on Managerial Decision-Making", Management Science, Vol. 15, p. 415, April 1969.

69. LaMotte, L. R., et al., Computational Efficiency in the Selection of Regression Variables, Themis TR#14, Texas A&M College Station, Texas, July 1969.

70. Madden, J. M., An Application to Job Evaluation of a Policy-Capturing Model for Analyzing Individual and Group Judgments, 6570th Personnel Research Laboratory, Aerospace Medical Division, Air Force Systems Command, PRL-TDR-63-15, May 1963.

71. Majone, Grandemenico, Classification and Discrimination in the Analysis of Credit Risks: I, Carnegie-Mellon University, Pittsburgh, February 1968.

72. Mead, D. F., Development of an Equation for Evaluating Job Difficulty, Report No. AFHRL-TR-70-42, AF Human Resources Laboratory, Lackland AFB, Texas, 19 pp., AD 720253, November 1970.

73. Mead, D. F., Continuation Study on Development of a Method for Evaluating Job Difficulty, Report No. AFHRL-TR-70-43, AF Human Resources Laboratory, Lackland AFB, Texas, 20 pp., AD 720254, November 1970.

74. Mead, D. F. and R. E. Christal, Development of a Constant Standard Weight Equation for Evaluating Job Difficulty, Report No. AFHRL-TR-70-44, AF Human Resources Laboratory, Lackland AFB, Texas 20 pp., AD 720255, November 1970.

75. Meehl, P. E., Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Literature, University of Minnesota Press, Minneapolis, 1954.

76. Mendenhall, W., Introduction to Linear Models and the Design and Analysis of Experiment, Wadsworth Publishing Co., 1968.

77. Milburn, T. W., et. al., The Echo Method and the Study of Values, Tech. Memo 951, General Research Corp., Santa Barbara, Calif., November, 1968.

78. Miller, George A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", Psychological Review, Vol. 63, p. 81-98, 1956.

79. Mood, A. M. and F. A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill, 1963.

80. Moore, Geoffrey, H. and Philip A. Klein, The Quality of Consumer Installment Credit, National Bureau of Economic Research, Columbia University Press, New York, 1967.

81. Morrison, D. F., _Multivariate Statistical Methods_, McGraw-Hill, New York, 1966.

82. Mullins, Cecil J. and Usdin, Eugene, _Estimation of Validity in the Absence of a Criterion_, AFHRL-TR-70-36, Personnel Division, AFHRL, Texas, 1970.

83. Myers, James H., "Numerical Scoring Systems for Retail Credit Evaluation", _Credit World_, Vol. 50, No. 7, p. 5-7, April, 1962.

84. Myers, James H. and Edward Forgy, "The Development of Numerical Credit Evaluation Systems", _American Statistical Association Journal_, Vol. 58, p. 799, 1963.

85. Naylor, J. C. and R. J. Wherry, Sr., _Feasibility of Distinguishing Supervisors' Policies in Evaluation of Subordinates by Using Rating of Simulated Job Incumbents,_ PRL-TR-64-25, Air Force Human Resources Laboratory, Lackland AFB, Texas, October 1964.

86. Naylor, J. C., and Schenck, E. A., "ρ_m As An Error Free Index of Rater Agreement", _Educational and Psychological Measurement_, Vol. 26, p. 815-824, 1966.

87. Norman, James E., David, H. A., "Restricted Ranking", _Psychometrics_, Vol. 34, No. 1, p. 85-110, March 1969.

88. Phillips, P. J., _An Evaluation of the Discriminant Assignment Model_, Master's Thesis, Air Force Institute of Technology, AD 806484, 59 pp., August 1966.

89. Roberts, Fred S., _What if Utility Functions Do Not Exist_, R-528-ARPA, RAND Corp., Santa Monica, Calif., September 1970.

90. Robinson, William O., _Multivariate Probability Assessment_, Technical Report No. 60, Operations Research Center, MIT, Cambridge, Massachusetts, 1971.

91. Rosenthal, R., Experimental Effects in Behavioral Research, Appleton-Century-Crofts, New York, 1966.

92. Sawyer, Jack, "Measurement and Prediction, Clinical and Statistical", Psychological Bulletin, Vol. 66, p. 178-200, 1966.

93. Searle, S. R., Linear Models, Wiley, New York, 1971.

94. Slovic, P., "Cue Consistency and Cue Utilization in Judgment", American Journal of Psychology, Vol. 79, p. 427-434, 1966.

95. Slovic, Paul, "Analyzing the Expert Judge: A Description Study of A Stockbroker's Decision Processes", Journal of Applied Psychology, Vol. 53, p. 255, 1969.

96. Slovic, Paul, Structural Determinants of Cue Utilization in Judgment, Oregon Research Institute, May 1969.

97. Slovic, P., and S. Lictenstein, "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment", ORI Monograph, Vol. 10, No. 1, July 1970.

98. Smith, R. L., et al., Acquisition of Restore Time Data by Subjective Techniques, AMRL-TR- 0-76, Aerospace Medical Research Laboratory, Wright Patterson AFB, Ohio, NTIS AD 727015, March 1971.

99. Smith, T. H., C. D. Gott, and R. A. Bottenberg, Predicting the Potential for Active Duty Success of Rehabilitated Air Force Prisoners, PRL-TR-67-16, Air Force Human Resources Laboratory, Lackland AFB, Texas, NTIS AD 672958, October 1967.

100. Sonquist, J. A. and J. N. Morgan, The Detection of Interaction Effects, Survey Research Center, University of Michigan, Ann Arbor, 1964.

101. Sonquist, J. A. "Finding Variables that Work", Public Opinion Quarterly, Vol. 33, No. 1, pp. 83-95, Spring 1969.

102. Sonquist, John A., Multivariate Model Building, Survey Research Center, University of Michigan, 1970.

103. Sonquist, J. A., E. L. Baker and J. N. Morgan, Searching for Structure, ISR, Ann Arbor, Michigan, 1971.

104. Stephenson, R. and Joe H. Ward, Jr., "Computer Assisted Decisions", Proceedings of the 78th Annual Convention, American Psychological Association, 1970.

105. Stimson, David H., "Utility Measurement in Public Health Decision", Management Science, Vol. 16, p. 17-31, October 1969.

106. Tatsuaka, M. M., "Joint Probability of Membership and Success in a Group: An Index Which Combines the Information from Discriminant and Regression Analysis as Applied to the Guidance Problem", Harvard Studies in Career Development, No. 6, Harvard University, Boxton, 1957.

107. Taylor, Bernard H., Groups of Homogeneous Criterion Variables in Multivariate Prediction, University of Washington, Seattle, January 1962.

108. Thames, James P., Integrated Personnel Management Systems, Quantified for Greater Efficiency, Master's Thesis, T-1694, Graduate School of Business Administration, The University of Texas at Austin, 1964.

109. Valenzi, Enzo R., Configural Cue Utilization in Human Decision Making, Ph. D. Dissertation, Bowling Green State University, August, 1970.

110. Vandell, R. F., "Management Evaluation in the Quantitative World", Harvard Business Review, Vol. 48, p. 83, January-February 1970.

111. Wagner, G. R. and L. L. Gooch, "Quantifying Human Judgments for more Meaningful Computer Decision Models", in process, 1972.

112. Waligore, Robert S., "Presentation to American Industrial Bankers Association, November 12, 1971," The Morris Planning Company of California, San Francisco.

113. Ward, J. H., Jr., An Application of Linear and Curvilinear Joint Functional Regression in Psychological Prediction, AFPTRC-TR-54-86, Air Force Personnel and Training Research Center, Lackland AFB, Texas, December 1954.

114. Ward, J. H., Jr., Use of a Decision Index in Assigning Air Force Personnel, WADC-TN-59-38, Personnel Research Laboratory, Lackland Air Force Base, Texas April 1959.

115. Ward, Joe H., Jr., and Marion E. Hook, A Hierarchical Grouping Procedure Applied to a Problem of Grouping Profiles, ASD-TN-61-55, Aeronautical Systems Division, Personnel Laboratory, Lackland Air Force Base, Texas, October 1961.

116. Ward, Joe H., Jr., "Comments on the Paramorphic Represen tation of Clinical Judgment", Psychological Bulletin, Vol. 59, p. 74-76, 1962.

117. Ward, J. H., "Multiple Linear Regression Models", Computer Applications in the Behavioral Sciences, H. Borko (ed.), Prentice-Hall, Englewood Cliff, p. 205-237, 1962.

118. Ward, Joe H., Jr., and Kathleen Davis, Teaching a Digital Computer to Assist in Making Decisions, PRL-TDR-63-16, 6570th Personnel Research Laboratory, Lackland Air Force Base, Texas 1963.

119. Ward, J. H., Comment on "The Partitioning of Variance in School Effects Studies", American Educational Research Journal, Vol. 6, No. 3, p. 467-474, 1969.

120. Ward, Joe H., Jr.,"Analyzing Judgment and Decision Processes", Working Paper, Southwest Educational Development Laboratory Octiber 1970.

121. Ward, J. H. Jr., and E. Jennings, Introduction to Linear Models, Prentice-Hall, Englewood Cliffs, New Jersey, (in press), 1972.

122. Weber, Donald C., "The Analysis of Experimental Designs Using Time-Sharing Computer Terminals", Paper presented at Joint Statistical Meeting, Colorado State University, August 23, 1971.

123. Wiggins, N. S., and Hoffman, P. J., "Three Models of Clinical Judgment", Journal of Abnormal Psychology, Vol. 73, p. 70-78, 1968.

124. Wilt, G. A., Jr. and J. M. Tierney Jr., "Progressive Risk Analysis Through Credit Scoring", Credit World, Vol. 56, No. 6, p. 10, March 1968.

125. Yntema, D. B., and Torgerson, W. S., "Man-computer Cooperation in Decisions Requiring Common Sense", IRE Transactions on the Profession-l Group on Human Factors in Electronics, HFE 2, 1961.

126. Yntema, D. B., and L. Klem, "Telling a Computer How to Evaluate Multi-dimensional Situation", IEEE Transactions on Human Factors in Electronics, HFE-6, p. 1, 1965.

127. Zaegel, Richard J., "A Point Rating System for Evaluating Customers", Credit World, Vol. 52, No. 1, p. 9-13, October 1963.

128. Zaegel, Richard J., "After 10 Years of Credit Scoring", Credit World, Vol. 59, No. 10, p. 14, July 1971.

129. Zyskind, G., et al., Linear Models, Statistical Information and Statistical Inference, Aerospace Research Laboratories ARL 71-0076, Wright Patterson Air Force Base, Ohio, May 1971.